



SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET

Petra Bago

**MODEL VIŠERAZINSKE PREZENTACIJE
TEKSTOVA STARIJIH HRVATSKIH
RJEČNIKA**

DOKTORSKI RAD

Mentori:

prof. dr. sc. Damir Boras
doc. dr. sc. Nikola Ljubešić

Zagreb, 2014.



UNIVERSITY OF ZAGREB
FACULTY OF HUMANITIES AND SOCIAL SCIENCES

Petra Bago

MULTILEVEL PRESENTATION MODEL OF OLD CROATIAN DICTIONARY TEXTS

DOCTORAL THESIS

Supervisors:
prof. dr. sc. Damir Boras
doc. dr. sc. Nikola Ljubešić

Zagreb, 2014.

Doktorska disertacija izrađena je na

Filozofskom fakultetu Sveučilišta u Zagrebu

Odsjek za informacijske i komunikacijske znanosti

Katedra za računalnu obradu jezika, leksikografiju i
enciklopedistiku

Povjerenstvo za ocjenu doktorske disertacije:

1. dr. sc. Nives Mikelić Preradović, izv. prof., Filozofski fakultet Sveučilišta u Zagrebu, predsjednik povjerenstva
2. dr. sc. Zdenko Jecić, docent, Leksikografski zavod *Miroslav Krleža*
3. dr. sc. Tomislava Lauc, izv. prof., Filozofski fakultet Sveučilišta u Zagrebu
4. dr. sc. Damir Boras, red. prof., Filozofski fakultet Sveučilišta u Zagrebu
5. dr. sc. Nikola Ljubešić, docent, Filozofski fakultet Sveučilišta u Zagrebu

Povjerenstvo za obranu doktorske disertacije:

1. dr. sc. Nives Mikelić Preradović, izv. prof., Filozofski fakultet Sveučilišta u Zagrebu, predsjednik povjerenstva
2. dr. sc. Zdenko Jecić, docent, Leksikografski zavod *Miroslav Krleža*
3. dr. sc. Tomislava Lauc, izv. prof., Filozofski fakultet Sveučilišta u Zagrebu
4. dr. sc. Damir Boras, red. prof., Filozofski fakultet Sveučilišta u Zagrebu
5. dr. sc. Nikola Ljubešić, docent, Filozofski fakultet Sveučilišta u Zagrebu

Datum obrane doktorske disertacije: 29.rujna 2014.

ZAHVALE

Rad posvećujem svojim roditeljima Carmen i Darinku bez čije podrške, strpljivosti i razumijevanja ovo ne bi bilo moguće. Hvala vam na svemu.

Zahvaljujem svojim mentorima prof. dr. sc. Damiru Borasu i doc. dr. sc. Nikoli Ljubešiću na poticanju i podršci pri izradi rada te prijateljskim savjetima u svim mogućim i nemogućim trenucima.

Zahvaljujem članovima komisije doc. dr. sc. Zdenku Jeciću, prof. dr. sc. Tomislavi Lauc i prof. dr. sc. Nives Mikelić Preradović na korisnim savjetima.

Zahvaljujem Marku, Martini, Maji, Ani, Ivi i *dedi* na zabavnim trenucima.

Zahvaljujem svim prijateljima na upornom ispitivanju o napretku rada, posebno Franciski i Ivi (Sv.) čija je potpora premašila prijateljske obaveze.

Petra Bago

SADRŽAJ

PREDGOVOR.....	10
1. Uvod.....	15
2. Cilj i hipoteze istraživanja.....	19
3. Osnovni podatci o odabranim rječnicima	21
3.1. Uvod.....	21
3.2. Faust Vrančić i njegov <i>Rječnik pet najuglednijih europskih rječnika</i>	23
3.2.1. Životopis Fausta Vrančića	23
3.2.2. <i>Dictionarium quinque nobilissimarum Europae linguarum Latinae, Italicae, Germanicae, Dalmatiae et Ungaricae</i>	25
3.3. Petr Loderecker i njegov <i>Dictionarium septem diversarum linguarum</i>	28
3.3.1. Životopis Petra Lodereckera.....	28
3.3.2. <i>Dictionarium septem diversarum linguarum videlicet Latine, Italice, Dalmatice, Bohemice, Polonice, Germanice et Ungarice</i>	29
3.4. Jakov Mikalja i njegov rječnik <i>Blago jezika slovinskoga</i>	30
3.4.1. Životopis Jakova Mikalje	30
3.4.2. <i>Blago jezika slovinskoga illi Slovník u Komu izgovarajuse rjeci slovinske Latinski i Diacki. Thesaurus linguae Illyricae sive Dictionarium Illyricum. In quo verba Illyrica Italice, et Latine redduntur</i>	32
3.5. Ardelio della Bella i njegov <i>Dizionario italiano-latino-illirico</i>	33
3.5.1. Životopis Ardelija della Belle	33
3.5.2. <i>Dizionario italiano-latino-illirico A cui si premettono alcune brevi Istruzioni Gramaticali,</i>	

	necessarie per apprendere la Lingua e l'Ortografia Illirica.....	36
3.6.	Joakim Stulli i njegovo <i>Rjecsosloxje ilirsko (slovinsko)- italiansko-latinsko</i>	41
3.6.1.	Životopis Joakima Stullija.....	41
3.6.2.	Rjecsosloxje ilirsko (slovinsko)-italiansko- latinsko u komu donosuse upotrebljenja, urednia, mucsnia istieh jezika krasnoslovja nacsini, izgovaranja i prorjecsja.....	42
3.7.	Ivan Mažuranić i Jakov Užarević te njihov <i>Njemačko- ilirski slovar</i>	43
3.7.1.	Životopis Ivana Mažuranića i Jakova Užarevića	43
3.7.2.	Deutsch-ilirisches Wörterbuch. Njemačko-ilirski slovar.....	45
3.8.	Mirko Divković i njegov <i>Latinsko-hrvatski rječnik za škole</i>	47
3.8.1.	Životopis Mirka Divkovića	47
3.8.2.	Latinsko-hrvatski rječnik za škole.....	49
3.9.	Dragutin Antun Parčić i njegov <i>Rječnik hrvatsko- talijanski</i>	51
3.9.1.	Životopis Dragutina Antuna Parčića	51
3.9.2.	Rječnik hrvatsko-talijanski. Vocabolario croato- italiano	53
4.	Digitalizacija rječnika	63
4.1.	Opća načela digitalizacije	63
4.2.	Svjetska i europska iskustva	69
4.2.1.	Projekt <i>Gutenberg</i>	70
4.2.2.	Projekt TITUS	75
4.2.3.	Projekt <i>Perseus</i>	80
4.3.	Hrvatska iskustva.....	85
4.3.1.	Arhivski informacijski sustav <i>ARHiNET</i> Hrvatskoga državnog arhiva.....	86
4.3.2.	Projekt <i>Digitalizirana baština</i> Nacionalne i sveučilišne knjižnice.....	90
4.3.3.	<i>Portal hrvatske rječničke baštine</i> projekta Hrvatska rječnička baština i prikaz rječničkoga znanja.....	93

4.4.	Zaključak.....	95
5.	Obrada prirodnoga jezika (OPJ) i povijesni tekstovi.....	96
5.1.	Uvod.....	96
5.1.1.	Digitalizacija kulturne baštine	97
5.1.2.	Obrada prirodnoga jezika i povijesni tekstovi.....	99
5.2.	Značajke suvremenih i povijesnih tekstova	101
5.2.1.	Značajke suvremenih tekstova	103
5.2.1.1.	Standardna inačica.....	103
5.2.1.2.	Standardni pravopis.....	104
5.2.1.3.	Tekstovi u digitalnom obliku	104
5.2.1.4.	Postojeći resursi i alati za obradu prirodnog jezika.....	104
5.2.2.	Značajke povijesnih tekstova.....	105
5.2.2.1.	Nedostatak standardne inačice	105
5.2.2.2.	Nedostatak standardnoga pravopisa	106
5.2.2.3.	Nedostupnost tekstova u digitalnom obliku	110
5.2.2.4.	Nepostojeći resursi i alati za obradu prirodnog jezika za povijesne tekstove	110
5.3.	Zaključak.....	111
6.	Prijedlog modela	113
6.1.	Uvod.....	113
6.2.	Model prikaza znanja u rječnicima.....	114
6.2.1	Prikaz sheme	120
6.2.1.1	Zaglavlje	120
6.2.1.2	Natuknice.....	145
6.2.2	Primjer natuknice cijele označene TEI oznakama.....	150
6.3.	Primjena poluautomatskih metoda za opis shemom.....	155
6.3.1.	Osnovni pojmovi strojnog učenja.....	155
6.3.2.	Skup podataka.....	160
6.3.2.1	Pojavnice	160
6.3.2.2	Skup za učenje.....	170
6.3.2.3	Primjer natuknice označene na jezičnoj i strukturnoj razini	174
6.3.3.	Metoda	181

6.3.3.1	Evaluacijske mjere	182
6.3.3.2	Faze istraživanja	183
6.3.3.3	n-struka unakrsna validacija.....	185
6.3.4.	Označavanje jezika	186
6.3.5.	Označavanje strukture	190
6.3.6.	Eksperiment s ručnim označavanjem i brzinom ispravaka	195
6.3.7.	Završni eksperiment na skupu za provjeru.....	197
6.3.8.	Krivulja učenja.....	201
6.4.	Povezivanje koncepata na resurse	203
6.5.	Znanstveni doprinos modela.....	204
6.6.	Mogućnosti daljnjih istraživanja.....	206
7.	Zaključak.....	208
8.	Literatura	219
9.	Sažetak	232
10.	Summary	233
PRILOZI	234
11.	Popis natuknica izdvojenih za ručno označavanje.....	235
11.1.	Izdvojene natuknice u obliku priređenom za obradu s vidljivim oznakama.....	235
11.2.	Izdvojene natuknice u izvornom obliku	265
12.	Popis kratica navedenih u rječniku i njihova tumačenja.....	285
12.1.	Popis autora i djela korišten kao korpus za izradu rječnika .	285
12.2.	Popis ostalih kratica	299
13.	Popis tablica	301
14.	Popis slika	302
15.	Životopis.....	303
15.1	Popis objavljenih radova.....	304
15.1.1	Poglavlja u knjizi	304
15.1.2	Znanstveni radovi u drugim časopisima.....	304
15.1.3	Znanstveni radovi u zbornicima skupova s međunarodnom recenzijom	305
15.1.4	Sažeci u zbornicima skupova	305
15.1.5	Neobjavljena sudjelovanja na skupovima	306
15.1.6	Druge vrste radova.....	306

PREDGOVOR

Razvoj informacijsko-komunikacijskih i jezičnih tehnologija omogućuje jednostavniji pristup i brži protok informacija, pohranu velikih količina podataka te sve jednostavnije rukovanje njima. Sve jednostavnija primjena tehnologija omogućuje njezino prodiranje u sve segmente ljudskog djelovanja, pa tako i u kulturnu djelatnost. Stoga je razvidno da se sve više projekata koji uključuju kulturnu baštinu okreću upravo navedenim tehnologijama kako bi omogućili pristup, protok, pohranu i rukovanje baštinom. Projekt „Hrvatska rječnička baština i hrvatski europski identitet“ glavnog voditelja prof. dr. sc. Damira Borasa bavi se istraživanjem hrvatske rječničke baštine.

Od 1595. godine kada je objavljen prvi samostalni tiskani hrvatski rječnik Fausta Vrančića pa do kraja 19. stoljeća u Hrvatskoj je objavljeno više od 300 različitih jednojezičnih, dvojezičnih, višejezičnih, enciklopedijskih, tehničkih, vojnih i drugih rječnika koji predstavljaju neprocjenljivu hrvatsku rječničku baštinu (od kojih je većina nedostupna za javnu upotrebu), a dio tih rječnika je strojno obrađen u sklopu projekta. „Taj velik broj rječnika objavljen u velikom broju europskih gradova (u prvim stoljećima uglavnom izvan hrvatskih prostora) potvrđuje pretpostavku da su Hrvatska i hrvatski jezik (u svim tadašnjim nazivima i idiomima) bili prisutni u Europi, da su bili neodvojivi

dio europskog identiteta te da su svojim sadržajem povratno utjecali na svijest o hrvatskom identitetu u Europi.“¹

Kao rezultat toga, ali i prethodnoga projekta istoga voditelja pod naslovom „Hrvatska rječnička baština i prikaz rječničkoga znanja“,² digitaliziran je velik broj rječnika od kojih je nekoliko, kao rezultat toga projekta, ogledno prikazano na Portalu hrvatske rječničke baštine.^{3,4}

Osim hrvatskih starijih rječnika digitaliziranih u sklopu ovih dvaju projekata i djelomično prikazanih na Portalu hrvatske rječničke baštine, mnogo digitalizirane građe objavljeno je i na mnogim drugim stranicama specijaliziranih ustanova za prikupljanje i pohranu povijesne građe poput arhiva, knjižnica i muzeja.

Pri tome je potrebno posebno spomenuti stranice Hrvatskog državnog arhiva i njegov informacijski sustav Arhinet koji predstavlja *online* registar arhivskih fondova i zbirki Republike Hrvatske te je središnja nacionalna evidencija arhivskoga gradiva koju vodi Hrvatski državni arhiv, a također, omogućuje i pristup sadržaju dokumenata pretraživanjem prema bilo kojoj riječi iz jedinice opisa ili kombiniranjem više kriterija iz jedinica opisa arhivskih jedinica, stvaratelja i imatelja. U Arhinetu je moguće i pregledati

¹**Boras, Damir** (voditelj). Hrvatska rječnička baština i hrvatski europski identitet; Znanstveni projekt, Zagreb, 2007-2013. http://zprojekti.mzos.hr/public/c2prikaz_det.asp?cid=1&psid=25&ID=451 (25.8. 2014.).

²**Boras, Damir** (voditelj). Hrvatska rječnička baština i prikaz rječničkoga znanja; Znanstveni projekt, Zagreb, 2003-2006. http://zprojekti.mzos.hr/public/c2prikaz_det.asp?cid=1&psid=25&ID=451 (25.8. 2014.).

³**Boras, Damir** (ur). Portal hrvatske rječničke baštine – Croatian Old Dictionary Portal, Zagreb, 2005, <http://crodip.ffzg.hr/>. (25. 8. 2014.).

⁴**Boras, Damir; Ljubešić, Nikola; Mikelić, Nives**. Croatian Old Dictionary Portal (CRODIP) // LIDA 2005 Proceedings: Libraries in the digital age 2005, Dubrovnik-Mljet, 2005., 127-137.

digitalne snimke najznačajnijih i najčešće korištenih dokumenata te dodatne informacijske izvore“.⁵

Također treba spomenuti i digitalnu građu Nacionalne i sveučilišne knjižnice (NSK). Osim kataloga NSK, hrvatskih bibliografija, biltena prinova, elektroničkih izvora i odabranih izvora podataka, naročito su zanimljive digitalne zbirke koje obuhvaćaju zvuke prošlosti, digitaliziranu baštinu, stare hrvatske novine, stare hrvatske časopise, hrvatski arhiv *weba*, digitalni akademski repozitorij, te virtualnu zbirku djela Ruđera Boškovića.

Nama je najzanimljivija tzv. digitalizirana baština⁶ koja obuhvaća stare knjige, grafičku zbirku, karte i planove, glazbenu građu, rukopise, tematske zbirke, novine i časopise te disertacije pri čemu digitalna zbirka starih knjiga za sada⁷ obuhvaća relativno malen broj djela hrvatskih autora, te samo jedan

⁵Arhinet, arhivski informacijski sustav. Hrvatski državni Arhiv. Zagreb. <http://arhinet.arhiv.hr/> (26. 8. 2014.).

⁶Projekt digitalizirane baštine Nacionalne i sveučilišne biblioteke, Zagreb, 2013. <http://db.nsk.hr/> (26. 8. 2014.).

⁷ Digitalna zbirka starih knjiga obuhvaća 19 djela hrvatskih autora i jedan zvučni zapis pjesme *Jur nijedna na svit vila* Hanibala Lucića u izvedbi dramskog umjetnika Joška Ševe. Pojedina su djela opremljena dodatnim kontekstualnim podacima. Obuhvaćena djela su: (1)*Glagoljički Misal po zakonu rimskoga dvora* kao prva hrvatska tiskana knjiga iz 1483.,(2) djela Marka Marulića –*Judita* (1521.); *De institutione bene vivendi* (1506.); *Epistola* (1522.),(3) djela Fausta Vrančića –*Machinae novae* (1615. – 1616.); *Rječnik pet najuglednijih europskih jezika* (1595.),(4) djela Ivana Gundulića –*Suze sina razmetnoga* (1622.); *Osman* (nastao 1628. – 1638., objavljen 1826.),(5) djelo Hanibala Lucića *Skladanja izvarsnih pisan razlicih* koja sadrži i dramu *Robinja* te zvučni zapis pjesme *Jur nijedna na svit vila* (1553.),(6) djelo Brne Karnarutića *Vazetje Sigeta grada* (1584.),(7) djelo Ivana Česmičkog [Jana Panonija] –*Ad Guarinum Veronesem Panegyricus* (1553.),(8) djelo Bartola Kašića *Institutionum linguae Illyricae libri duo* (1604.),(9) djelo Markantuna de Dominisa *De radiis visus et lucis in vitris perspectivis et iride* (1611.),(10) djelo Ivana Lučića *De regno Dalmatiae et Croatiae libri sex* (1666.),(11) djelo Pavla Rittera-Vitezovića *Croatia rediviva* (1700.),(12) djela Đure Baglivija *Opera omnia medico-practica et anatomica* (1710.),(13) djelo Matije Antuna Relkovića *Satyr iliti Divji csovik u vershe Slavoncem* (1762.),(14) djelo Ljudevita Gaja *Kratka osnova horvatsko-slavenskoga pravopisa* (1830.) te (15)

rječnik (Vrančičev) koji je ionako već digitaliziran u okviru projekta „Hrvatska rječnička baština i prikaz rječničkoga znanja“ te prikazan na Portalu hrvatske rječničke baštine. Ipak, i ta zbirka, iako za sada ograničena i većim dijelom svojim sadržajem izvan konteksta ovoga rada, predstavlja značajan resurs na kojem će se moći primijeniti tehnike i tehnologije označavanja rječničkih i povijesnih tekstova razvijenih i prikazanih u ovoj disertaciji.

S druge strane i u Europi i u svijetu pojavilo se mnogo izdanja digitaliziranih rječnika raznih jezika u rasponu od prvih tiskanih, ali i rukopisnih rječnika sve do danas, te s velikim brojem jezika, pri čemu ima malo znanstvene literature koja bi teoretski osmislila načine na koje treba prikazati te digitalizirane rječnike.

Pri tome nije samo riječ o internetskom prikazu i mogućnostima pretraživanja nego i o kvaliteti arhiviranja digitaliziranih rječnika te raznim drugim parametrima koji svi zajedno kao cjelina omogućuju kvalitetan pristup ne samo tekstovima starih rječnika nego i načinu njihove pripreme u grafičkom i strukturnom smislu.

S obzirom na to da većina tih stranica omogućuje pristup digitaliziranoj baštini uglavnom u slikovnom obliku, a tek u manjoj mjeri digitaliziranim tekstovima u cijelosti, te budući da su projekti hrvatske rječničke baštine u cijelosti orijentirani na prikaz (uz slikovni) u obliku digitalnog teksta, u ovom ćemo se radu prvenstveno usmjeriti na prikaz i označavanje rječničkih tekstova i prepoznavanje njihove strukture, jer je on najmanje sustavno proučavan segment

djela iz Knjižnice obitelji Zrinski (Bibliotheca Zriniana) koja sadrži preko 500 knjiga, među njima i 5 inkunabula, a za sada je dostupno samo jedno djelo *Augustissimorum Imperatorum* (1601.) koje sadrži bakroreze/portrete istaknutih careva i uglednika koji je uredio austrijski diplomat Jakob Schrenk von Notzing. <http://db.nsk.hr/HeritageDetails.aspx?id=1193> (28. 8. 2014.).

digitalizacije i jer je model prikaza tekstova važan za projekt „Hrvatska rječnička baština i hrvatski europski identitet“ kako bi se mogli bolje publicirati tekstovi rječnika digitaliziranih u projektu.

Stoga je cilj istraživanja ove disertacije, a u sklopu tog projekta, razvoj modela koji će omogućiti višerazinski prikaz tekstova starijih hrvatskih rječnika. Uz to je cilj i razviti automatske i poluautomatske metode za obradu digitaliziranih povijesnih tekstova koji sadrže različita dijakronijska stanja jezika.

1. Uvod

Sve veća dostupnost digitaliziranih povijesnih tekstova ima za posljedicu sve veće zanimanje za primjenu metoda, a time i alata, za obradu prirodnoga jezika (OPJ) na takvoj vrsti tekstova. Međutim, povijesni tekstovi imaju nekoliko značajki koje predstavljaju prepreke za primjenu spomenutih alata u onom obliku u kakvom trenutno postoje, s obzirom na to da se temelje na suvremenim jezicima.

Prirodni jezik možemo definirati kao jezik kojim se govori i koji je nekome materinski.⁸ Naspram prirodnoga, postoje i formalni jezici koje oblikuju ljudi kako bi jednoznačno komunicirali s računalima. Budući da je danas većina informacija zapisana prirodnim jezikom, područje koje se bavi problemom obrade podataka kodiranih prirodnim jezikom naziva se obrada prirodnog jezika. Višeznačnost prirodnog jezika osnovna je razlika između prirodnih i formalnih jezika. Stoga je temeljni napor u području obrade prirodnog jezika formalizacija prirodnog jezika bez gubitaka informacije da bi on bio strojno obradiv.⁹

Povijesni tekstovi sadrže prirodni jezik, baš kao što i suvremeni tekstovi sadrže jezike kojima danas govorimo i pišemo. Prema tome, neki problemi povijesnih tekstova isti su kao i kod suvremenih (npr. homonimija i polisemija).

⁸ Prema natuknici "Prirodni Jezik.", *Hrvatski Jezični Portal*. Novi Liber. http://hjp.noviliber.hr/index.php?show=search_by_id&id=fVZvWhc%3D&keyword=jezik. (2. 2. 2014.).

⁹ **Ljubešić, Nikola.** "Pronalaženje događaja u višestrukim izvorima informacija." Doktorska disertacija. Filozofski fakultet, Sveučilište u Zagrebu, 2009.

Međutim, za potrebe ovog istraživanja zanimaju nas razlike između suvremenih i povijesnih tekstova.

Suvremeni tekstovi sadrže suvremene jezike koji imaju svoju standardnu inačicu. To su uglavnom jezici koji se poučavaju u osnovnim i srednjim školama i školama stranih jezika, a obilježava ih dobra dokumentiranost u rječnicima i gramatikama. Osim standardne inačice, suvremeni jezici imaju standardni pravopis kojeg se pridržava većina objavljenih tekstova. Autori tekstova mogu u rječnicima provjeriti kako se određena riječ pravilno piše. U većini slučajeva, alati za obradu prirodnoga jezika, koji se temelje na pravilima ili statistici, mogu računati na to da svaka leksička jedinica ima malen i konačan broj oblika koji se mogu pojaviti u tekstovima. Osim toga, standardizirani pravopis omogućava ispravak nepravilno napisanih riječi.¹⁰

Metode, a time i alati, za obradu prirodnoga jezika nailaze na prepreke kada se primjenjuju na povijesne tekstove.¹¹ Tekstovi starijih hrvatskih rječnika imaju dvije važne značajke koje ih razlikuju od suvremenih. Prvo, hrvatski jezik u tekstovima obrađenih rječnika nije bio standardiziran u današnjem smislu. Jezik u tim rječnicima jest književni, jer se dobrim dijelom oslanjao na izvore odnosno djela hrvatskih pisaca (npr. della Bellin), ali nije bio normiran ni u jezičnom ni u pravopisnom smislu. Hrvatski je jezik normiran u obliku kakav ima danas tek potkraj 19. i početkom 20. stoljeća, prateći tijek događaja u Europi. To je razdoblje kada se javila zamisao da se književni jezik temelji na štokavskome i jekavskome, odbacivši kajkavsko i čakavsko narječje kao standarde. U isto vrijeme kada se normirao hrvatski književni jezik, ustaljen je i

¹⁰Piotrowski, Michael. 2012. *Natural Language Processing for Historical Texts*. Ed. Graeme Hirst. Synthesis Lectures on Human Language Technologies 17. Morgan & Claypool Publishers. <http://www.morganclaypool.com/doi/abs/10.2200/S00436ED1V01Y201207HLT017>. (str. 2).

¹¹Ibid. (str. 3).

današnji pravopis, a do tada nije bilo standardiziranog pravopisa. Stariji hrvatski rječnici obuhvaćaju i kajkavsko i čakavsko narječje, a pravopisno su se oslanjali na prakse koje su postojale u talijanskom, njemačkom ili mađarskom jeziku. Stoga se jedna riječ pisala na više načina, ovisno o autorovim sklonostima. Na primjer, „odreći se“ Jakov Mikalja sredinom 17. st. zapisuje u obliku „odreccife“, Joakim Stulli početkom 19. st. „odrèchi se“, a Mirko Divković i Dragutin Antun Parčić početkom 20. st. bilježe „odreći se“. Nestandardizirani je pravopis velika prepreka obradi prirodnoga jezika jer se konvencionalne metode i tehnike oslanjaju na pretpostavku da su tekstovi pisani na jeziku sa standardnim pravopisom. U suprotnom statističke metode nailaze na prepreke jer dolazi do većeg stupnja nesigurnosti nego u standardiziranom jeziku. Dodatna značajka povijesnih tekstova jest ta što se pravopis može razlikovati između dokumenata iz istog razdoblja; dapače, jedan autor u jednom tekstu može napisati riječ na više načina.

Inačice pisanja također postoje u suvremenim jezicima, ne samo u povijesnim. Na primjer, postoje britanski i američki pravopisi (brit. engl. „colour“ naspram amer. engl. „color“) ili njemački i švicarski pravopisi (njem. „die Straße“ naspram švic. „die Strasse“). Međutim, različite inačice pisanja znatno su manje u suvremenim tekstovima nego u povijesnim. Osim toga, iz navedenih primjera vidljivo je da se radi o standardima u Velikoj Britaniji i Sjedinjenim Američkim Državama, odnosno Njemačkoj i Švicarskoj, što znači da razlike u načinu pisanja ustvari podliježu pravilima te stoga ne predstavljaju problem alatima za obradu prirodnoga jezika, ako se oni prilagode određenoj inačici tog jezika. Doduše, postoje okolnosti kod suvremenih jezika kada se pojavljuju različite inačice pisanja koje ne podliježu pravilima, a to su pravopisne pogreške, što predstavlja značajan problem kod pretraživanja teksta. Osim toga, kod suvremenih jezika problem predstavljaju nepoznate, ali pravilno

napisane riječi, kao što su neologizmi. Iako mrežne tražilice danas mogu rukovati određenim brojem inačica kako bi izašle na kraj s pogreškama u tipkanju, one i dalje pretpostavljaju da postoji samo jedan ispravan oblik za svaku riječ.

Uz sve prethodno navedeno, razumljivo je da ne postoje razvijeni alati za obradu prirodnog jezika za povijesne tekstove na hrvatskom. Stoga je cilj istraživanja postaviti temelje koji će omogućiti daljnji razvoj resursa i alata za obradu hrvatskoga jezika u povijesnim tekstovima. Razvit će se metode za automatsku i poluautomatsku obradu digitaliziranih povijesnih tekstova, ispitati njihova točnost i testirati koliko automatska ili poluautomatska obrada ubrzava proces naknadnog označavanja, odnosno dovođenja teksta do prihvatljive razine točnosti označavanja. Koristeći navedene metode, naredni je cilj razviti model višerazinskog prikaza tekstova starijih hrvatskih rječnika.

2. Cilj i hipoteze istraživanja

U posljednjih nekoliko desetljeća u području obrade prirodnoga jezika razvijeno je mnogo jezičnih resursa, jezičnih alata i ostalih kompleksnijih sustava za obradu prirodnoga jezika. Međutim, razvoj se odvijao izolirano na projektima, sveučilištima i institutima, ali i od strane pojedinaca, gdje je nedostajalo komunikacije među članovima zajednice. Rezultat toga su resursi, alati i sustavi koji ne mogu međusobno razmjenjivati informacije.

Stoga je cilj istraživanja ovog rada razvoj modela koji će omogućiti prikaz znanja u tekstovima starijih hrvatskih rječnika. Rječnici sadrže svome dobu suvremeni jezik koji se razvijao tijekom stoljeća, te su u svakome prikazana različita dijakronijska stanja jezika, vremenski i pravopisno omeđena. Sedam starih hrvatskih rječnika obuhvaćenih ovim radom sadrže inačice hrvatskoga jezika u rasponu od preko 300 godina: od 1595. do 1901.

Model će zadržati vjerodostojnost i izvornost prikaza svakog pojedinog jezika te definirati shemu za označavanje digitaliziranih tekstova rječnika u skladu s najnovijim *de facto* standardima (*Text Encoding Initiative*, TEI), a time će se dobiti na interoperabilnosti podataka. Osim toga, služiti će kao temelj za izgradnju računalnih sustava primjenjivih na digitaliziranu hrvatsku rječničku baštinu. Razvojem sustava potrebno je osigurati vezu na današnje domaće suvremene (digitalizirane) rječnike (na primjer Hrvatski jezični portal) i enciklopedije (na primjer Hrvatska enciklopedija Leksikografskog zavoda Miroslav Krleža, Proleksis enciklopedija), ali i na međunarodne projekte prikaza

i obrade jezika sadržanih u odabranim rječnicima (latinski, talijanski, njemački, mađarski, češki, poljski). Zahvaljujući formaliziranom prikazu rječničkih natuknica putem TEI oznaka, moguće je prikazati koncepte unutar rječnika putem RDF-a (*Resource Description Framework*). Tako definirani koncepti mogu se naknadno povezati na resurse nastale projektima poput Linked Data i DBpedije, čime će se omogućiti istraživanje odnosa i svojstava rječničkog teksta.

Osim toga, uvidjelo se da automatske metode za opis prema nekoj zadanoj shemi imaju svoja ograničenja i domete te je jasno da nikad nije moguće razviti algoritam koji će potpuno točno označiti svaki tekst. Stoga je neophodan dodatni ljudski napor i vrijeme da se ti tekstovi dovedu do potpune točnosti (ili barem prihvatljive prema nekim statističkim pretpostavkama), kako bi bili prikladni za kasnije analize i obrade. Iz navedenog proizlazi sljedeći cilj ovog rada, a to je razviti metode za automatsku i poluautomatsku obradu digitaliziranih povijesnih tekstova, ispitati njihovu točnost i testirati koliko automatska ili poluautomatska obrada ubrzava proces naknadnog označavanja, odnosno dovođenja teksta do prihvatljive razine točnosti označavanja.

Ovim radom provjeravaju se dvije hipoteze:

1. S obzirom na stanje jezičnih tehnologija moguće je razviti metode za automatsku i poluautomatsku obradu povijesnih tekstova koji sadrže vremenski nezavisne inačice jezika.

2. Zahvaljujući tim metodama, moguće je izgraditi sveobuhvatni model prikaza znanja u starijim hrvatskim rječnicima.

3. Osnovni podatci o odabranim rječnicima

3.1. Uvod

Planirano istraživanje izvršit će se u sklopu znanstvenog projekta predloženog prvog mentora prof. dr. sc. Damira Borasa „Hrvatska rječnička baština i hrvatski europski identitet“ (2007. – 2013.). Projekt je financiran od MZOS-a, a nastavlja se na projekte „Hrvatska rječnička baština i prikaz rječničkog znanja“ (2002. – 2006., financiranog od MZOS-a) te „Croatian dictionary heritage“ (2003. – 2005., financiranog od UNESCO-a).

Za obradu, analizu i primjenu modela odabrani su sljedeći (kronološki navedeni) rječnici, svi među najznačajnijim djelima hrvatske leksikografije:

1. Lodereker, Petr. **Dictionarium septem diversarum linguarum videlicet Latine, Italice, Dalmatice, Bohemice, Polonice, Germanice et Ungarice**, Prag, 1605.
2. Mikalja, Jakov. **Blago jezika slovinskoga illi Slovník u Komu izgovarajuse rjeci slovinske Latinski i Diacki. Thesaurus linguae Illyricae sive Dictionarium Illyricum. In quo verba Illyrica Italice, et Latine redduntur**, Loreto/Ancona, 1649./1651.
3. della Bella, Ardelio. **Dizionario italiano-latino-illirico A cui si premettono alcune brevi Istruzioni Gramaticali, necessarie per**

apprendere la Lingua e l'Ortografia Illirica, 2. izdanje, Dubrovnik, 1785.

4. Stulli, Joakim. **Rjecsosloxje ilirsko (slovinsko)-italiansko-latinsko u komu donosuse upotrebljenja, urednia, mucsnia istieh jezika krasnoslovja nacsini, izgovaranja i prorjecsja**, I i II, Dubrovnik, 1806.
5. Mažuranić, Ivan i Jakov Užarević. **Deutsch-ilirisches Wörterbuch. Njemačko-ilirski slovar**, Verlag und Druck der k. priv. Ilirischen National-Buchdruckerei von Dr. Ljudewit Gaj. Zagreb, 1842.
6. Divković, Mirko. **Latinsko-hrvatski rječnik za škole**, Zagreb, 1900.
7. Parčić, Dragutin. **Rječnik hrvatsko-talijanski. Vocabolario croato-italiano**, 3. izdanje, Zadar, 1901.

Lodereckerov se rječnik temelji na Vrančićevom rječniku iz 1595., prvom hrvatskom tiskanom rječniku. Budući da je već izvršeno istraživanje razlika između Vrančićeva i Lodereckerova rječnika na projektu „Hrvatska rječnička baština i hrvatski europski identitet“¹², uzet je Lodereckerov rječnik, jer sadrži dodatna dva jezika (češki i poljski), a od Vrančićeva se razlikuje samo u 18 dodanih natuknica. Stoga u popis odabranih rječnika možemo dodati još i Vrančićev rječnik:

¹²Bago, Petra; Kolarec, Martina. **Računalna priprema, obrada i prikaz sedmojezičnog rječnika Petera Lodereckera povodom 400. godišnjice izdavanja**. Rektorova nagrada za studentski znanstveni rad. Filozofski fakultet, Sveučilište u Zagrebu, 2006.

8. Vrančić, Faust. **Dictionarium quinque nobilissimarum Europae linguarum Latinae, Italicae, Germanicae, Dalmatiae et Ungaricae.** Venecija, 1595.

U ovom poglavlju ukratko ćemo prikazati životopise leksikografa i prikazati osnovne podatke o odabranim rječnicima.

3.2. Faust Vrančić i njegov *Rječnik pet najuglednijih europskih rječnika*

3.2.1. Životopis Fausta Vrančića

Faust Vrančić (Šibenik, 1551. – Rim, 1617.),^{13,14} hrvatski je humanist, polihistor, izumitelj, filozof, pisac, povjesničar i leksikograf. Bio je tipičan predstavnik svojeg doba koji se isticao znanjem jezika, obrazovanošću te visokim položajima u crkvenoj hijerarhiji, na sveučilištima, u vojnoj službi i na kraljevskom dvoru. Poznao je barem sedam europskih jezika. Od njegovih tiskanih djela najznačajniji su *Dictionarium quinque nobilissimarum Europae linguarum, Latinae, Italicae, Germanicae, Dalmaticae et Ungaricae*, prvi hrvatski samostalni rječnik, te *Machinae novae Fausti Verantii Siceni cum declaratione Latina Italica Hispanica Gallica et Germanica*, jedno od najvažnijih djela iz područja tehnike s početka 17. st. u kojem se nalazi njegov *leteći čovjek*. Na hrvatskom jeziku je 1606. godine u Rimu objavio *Život*

¹³Bago, Petra; Kolarec, Martina. **Računalna priprema, obrada i prikaz sedmojezičnog rječnika Petera Lodreckera povodom 400. godišnjice izdavanja.** Rektorova nagrada za studentski znanstveni rad. Filozofski fakultet, Sveučilište u Zagrebu, 2006.

¹⁴Ravlić, Slaven, ed. "Vrančić, Faust." *Hrvatska Enciklopedija, Mrežno Izdanje*. Zagreb: Leksikografski zavod Miroslav Krleža, 2013. <http://www.enciklopedija.hr/Natuknica.aspx?ID=65386> (31. 8. 2014.).

nikoliko izabranih divic, zbornik hagiografija, prvenstveno svetica koje su posebno štovane u hrvatskim krajevima uz jadransku obalu. Djelo je posvetio opaticama šibenskoga samostana Sv. Spasa.

Faust Vrančić rođen je u uglednoj šibenskoj obitelji 1. siječnja 1551. godine. Osnovno obrazovanje stekao je u Požunu kod strica Antuna Vrančića koji je bio crkveni vjerodostojnik, diplomat i pisac. Domovinu je napustio radi daljnjeg školovanja. Većinu svog života proveo je u inozemstvu, najviše u Italiji, Mađarskoj i Austriji, što se kasnije odrazilo na njegovo jezično blago i njegovu ljubav prema hrvatskom jeziku. Zahvaljujući stričevoj funkciji ostrogonskog nadbiskupa, Vrančić je otišao na studij u Padovu i Veneciju, gdje je pokazao osobito zanimanje za filozofiju, fiziku i matematiku. Nakon studija 1571. godine nakratko se vratio u rodni Šibenik, ali je nedugo zatim otišao stricu u Ugarsku. Godine 1578. oženio se Marietom, a već je sljedeće godine postavljen za upravitelja biskupskih dobara u Veszprému u Ugarskoj. Za tajnika rimsko-njemačkog cara i hrvatsko-ugarskog kralja Rudolfa II. imenovan je 1581. Tu je bio okružen mnogim znanstvenicima i poznatim umjetnicima tog vremena te se u kraljevoj knjižnici mogao više posvetiti istraživanju problema prirodnih znanosti i tehnike. Tamo je boravio sve do 1594. godine kada je napustio dvor te otišao živjeti u Šibenik i Italiju. Kralj ga je 1598. godine imenovao naslovnim čanadskim biskupom u Mađarskoj, ali kako je ugarski Csanad tada bio pod osmanlijskom vlašću, bila je to tek čast koja je potvrđivala i isticala tajnikov ugled. Na njegovu žalost, supruga mu je umrla mlada ostavivši mu kćer Albu Rozu i sina, koji je umro još kao dječak. Nakon njezine smrti zaredio se 1600. godine te postao kraljevskim tajnikom za poslove Ugarske i Transilvanije. Dvor je napustio 1605., nezadovoljan državnom službom. Ušao je u Kongregaciju sv. Pavla u Rimu u red barnabita, koji su se tada angažirali u odgoju mladeži. Godine 1608. se odrekao biskupske časti. Ostatak života djeluje u Rimu. Nakon

što je teško obolio, uputio se prema domovini, ali je u Veneciji umro 20. siječnja 1617. godine. Prema oporučnoj želji njegovo je tijelo sahranjeno u domovini, u crkvi sv. Marije od Milosti u Prvić Luci na otoku Prviću kraj Šibenika.

Razni su narodi svojatali Vrančićevo podrijetlo, pa je tako bio Talijan i Mađar. Osim toga, njegov padobran (*homo volans*, tj. *leteći čovjek*) pripisan je Leonardu da Vinciju. Međutim, Vrančić je u knjizi *Novi strojevi* iznio koju državu smatra domovinom u opisu šibenske katedrale: „...iznimno je lijepa i neobična oblika, pa sam je htio, kao ures svoje domovine, uvrstiti ovamo među svoja nova otkrića...“.

3.2.2. Dictionarium quinque nobilissimarum Europae linguarum Latinae, Italicae, Germanicae, Dalmatiae et Ungaricae

Faust Vrančić je 1595. godine tiskao u Veneciji kod Nikole Morettija rječnik pod nazivom *Dictionarium quinque nobilissimarum Europae linguarum, Latinae, Italicae, Germanicae, Dalmaticae et Ungaricae* (*Rječnik pet najuglednijih europskih jezika, latinskoga, talijanskoga, njemačkoga, hrvatskoga i mađarskoga*). To je petojezični rječnik, prvi samostalno tiskani hrvatski višejezični rječnik, ali i prvi oveći mađarski rječnik. Ima 6 predgovornih i 128 rječničkih stranica, koje su organizirane u pet stupaca. Paginacija započinje tek rječničkim tekstom. Riječi su raspoređene u pet stupaca, a polazišni jezik je latinski. Zatim je riječi preveo na talijanski, njemački, hrvatski i mađarski jezik. Pretpostavlja se da je Vrančić koristio neki od već postojećih rječnika, vjerojatno drugo izdanje Calepinusova latinskog

rječnika iz 1564. godine, od kuda je preuzeo latinske natuknice i njihove prijevode na talijanski i njemački jezik.¹⁵

U Europi se pokazalo veliko zanimanje za Vrančićev rječnik te su ga mnogi leksikografi upotrebljavali pri sastavljanju svojih rječnika. Početkom 17. stoljeća njime su se poslužila i tri velika europska leksikografa. Među njima je bio Talijan Bernardin Baldi, koji je sastavio rukopisni mađarsko-talijanski rječnik. Zatim tu je bio i Hieronymus Megiser, njemački humanist, koji je 1603. sastavio prvi svejezični rječnik (*Thesaurus polyglottus*).¹⁶

Konačno, nama ovdje najvažniji, Petr Loderecker, koji je objavio drugo dopunjeno izdanje samo deset godina kasnije, 1605., pod naslovom *Dictionarium septem diversarum linguarum, videlicet Latine, Italice, Dalmatice, Bohemice, Polonice, Germanice et Ungarice* (Rječnik sedam različitih jezika, i to latinskog, talijanskog, dalmatinskog, češkog, poljskog, njemačkog i mađarskog). Dodao je češki i poljski. Treće izdanje objavio je 1834. godine mađarski leksikograf J. Török u Bratislavi, tadašnjem Požunu, pod naslovom *Dictionarium pentaglottum*, gdje je istaknuo važnost rječnika za mađarsku leksikografiju. U novije vrijeme rječnik doživljava tri dodatna izdanja izdana u Zagrebu. Izlaze 1971., 1990. i 1992. godine. Izdanja iz 1971. i 1992. sadrže pogovor prof. Ljudevita Jonkea, te *Hrvatsko-latinski rječnik 1595* kojeg je prema izvorniku Fausta Vrančića postupkom obrata izradio dr. Valentin Putanec.

¹⁵Lisac, Josip. Držićev jezik i leksikografija njegova doba // Čakavska rič. 35, 1(2007) str. 9-14.

¹⁶Rječnik sadrži popis riječi s njihovim ekvivalentima na drugom jeziku. Riječi su poredane abecednim redom prema lijevom, latinskom stupcu. Ekvivalenti su prevedeni na arapski, armenski, baskijski, staroslavenski, hrvatski, češki, danski, nizozemski, engleski, finski, francuski, njemački, grčki, hebrejski, mađarski, perzijski, poljski, slovenski, lužičkosrpski, španjolski, tatarski, turski i velški.

Od domaćih leksikografa, Vrančićevim rječnikom su se poslužili Jakov Mikalja, Juraj Habdelić, Ivan Belostenec te Ardelio della Bella. Do danas je izašlo osam izdanja Vrančićeva rječnika.

Vrančić je rječnik napisao leksikografski ispravno, tako da su imenice u nominativu, a glagoli u infinitivu. Pridjevi se nalaze u nominativu muškog roda, s nastavcima za ženski i srednji rod. Kao i većina srednjovjekovnih i prvih novovjekovnih rječnika, tako i ovaj rječnik ustvari sadrži popis riječi s njihovim ekvivalentima na drugim jezicima. Ukupno ima 5 411 natuknica abecedno poredanih prema latinskom, lijevom stupcu. Hrvatskih natuknica je 5 373, jer za 38 riječi nema hrvatskog prijevoda (navedenih u popisu niže s prijevodima na današnji hrvatski jezik). U talijanskom dijelu nedostaje prijevod za jednu riječ, u njemačkom nedostaju dvije riječi, a u mađarskom nijedna. Iz ovoga je vidljivo da je hrvatski očito Vrančiću bio najmanje poznat jezik, što i nije iznenađujuće s obzirom na činjenicu da je većinu vremena izbivao iz domovine.

Popis latinskih riječi bez hrvatskog prijevoda: Adulterare, Atrocitas, Carduelis, Confutare, Contumacia, Lena, Monedula, Monile, October, Pertinacia, Picus, Pila, Podium, Poena, Præsentia, Promptuarium, Pulegium, Pumilio, Renes, Sandalium, Secta, Securitas, Sententia, September, Sextilis, Sporta, Spurius, Stupa, Subula, Supinus, a, m, Talpa, Taxus, Temperantia, Tum, Vestibulum, Victor, Vlmus, Vpupa.

Popis latinskih riječi bez talijanskog prijevoda: Mufcus

Popis latinskih riječi bez njemačkog prijevoda: Crepare, Meatus

Rječnik sadrži popis slavenizama u mađarskom jeziku, što se može smatrati početkom etimološkoga rječnika. Većina leksika je čakavska, ali ima i štokavskoga, poneka i kajkavska riječ, pa i crkvenoslavenska.

3.3. Petr Loderecker i njegov *Dictionarium septem diversarum linguarum*

3.3.1. Životopis Petra Lodereckera

O životu Petra Lodereckera (Petrus Lodereckerus)¹⁷ (15./16. st) zna se vrlo malo. U enciklopedijama se o njemu može naići tek na neki usputni podatak, i to jedino u natuknicama o Vrančiću ili njegovom rječniku ili o slavistici. Svi podaci o životu Petra Lodereckera preuzeti su iz članka Radoslava Katičića *Lodereckerov rječnik – leksikografsko djelo i kulturni spomenik*. Loderecker je rođen u mjestu Prošovice, koje je danas dio grada Jablonec nad Nisou u Češkoj. Nije sigurno je li bio Čeh ili Nijemac, budući da su svi stanovnici tog mjesta početkom 20. stoljeća bili Nijemci, a nedostaju podaci s kraja 16. stoljeća. Također se ne zna kada je Loderecker pristupio redu sv. Benedikta u Pragu. Postoje podaci da je bio prior samostana u Broumovu, dok je prema drugoj vijesti bio prior u mjestu Kladruby u zapadnoj Češkoj. Nakon toga se pridružio benediktinskoj opatiji u Břevnovu, selu koje se nalazilo zapadno od Praga, a danas je njegovo predgrađe. Opat u samostanu Emaus u praškom Novom gradu postao je 1606. godine, a postavio ga je Wolfgang Selender, njegov zemljak iz Prošovice. Kad je u grad prodrila njemačka vojska iz Passaua i bila prisiljena povući se, dva su vojnika potražila utočište u samostanu gdje je Loderecker bio opat. Razjareni puk tada je provalio u samostan i poharao ga. Opat je preživio jer se uspio sakriti u samostanski dimnjak. Loderecker je bio i opat samostana Na Slovanech, no u tome nije bio uspješan, tako da je 1611. godine, u vrijeme teških vjerskih sukoba, neki vizitator (vjerojatno Wolfgang

¹⁷Bago, Petra; Kolarec, Martina. **Računalna priprema, obrada i prikaz sedmojezičnog rječnika Petera Lodereckera povodom 400. godišnjice izdavanja.** Rektorova nagrada za studentski znanstveni rad. Filozofski fakultet, Sveučilište u Zagrebu, 2006.

Selender) od njega zatražio da odstupi s funkcije. Još je poznat i podatak da je bio župnik u mjestu Zdice kraj Berouna. Umro je 30. prosinca 1636. godine u Pragu.

3.3.2. Dictionarium septem diversarum linguarum videlicet Latine, Italice, Dalmatice, Bohemice, Polonice, Germanice et Ungarice

Petr Loderecker je 1605. godine u Pragu objavio *Dictionarium septem diversarum linguarum, videlicet Latine, Italice, Dalmatice, Bohemice, Polonice, Germanice et Ungarice* (Rječnik sedam različitih jezika, odnosno engleski, talijanski, hrvatski, češki, poljski, njemački i mađarski). Rječnik se temelji na petojezičnom rječniku Fausta Vrančića iz 1595. godine, koji je Loderecker dopunio češkim i poljskim jezikom. Time je stvorena poveznica cijele srednje Europe. Knjiga ima 578 stranica, nema paginacije. Format stranice je položena osminka, odnosno albumski oblik tzv. quartformat, a odabran je kako bi na stranicu stalo sedam okomitih stupaca.

Kao što se malo podataka zna o Lodereckerovom životu, tako je i s njegovim radom na sedmojezičnom rječniku. Sve što znamo je tek ono što možemo saznati iz samog djela. Pošto Loderecker sam nije znao hrvatski, morao je imati suradnika koji mu je pomagao za hrvatski dio rječnika, a pretpostavlja se da su bila trojica. Jedan je hrvatski jezik znao izvrsno i napisao je na njemu predgovor u prvom dijelu. Vjerojatno je taj suradnik bio upravo Vrančić. Drugi je suradnik bio nesiguran u znanju hrvatskog jezika i napisao je predgovor za hrvatsko-latinski rječnik u drugom dijelu. Pretpostavlja se da je bio Hrvat koji je već dugo živio u Pragu ili Čeh koji je naučio ponešto hrvatskog. Treći suradnik uopće nije znao hrvatski i provodio je otiskivanje cijeloga rječnika.

Lodereckerovo izdanje rječnika je dvodijelno. Prvi dio je sedmojezični rječnik s latinskim kao polaznim jezikom i abecedno poredanim latinskim rječnicima, a zatim slijede jezici poredani slijedom iz naslova djela: talijanski, hrvatski, češki, poljski, njemački, mađarski. Taj dio ima 200 stranica i tiskao ga je praški tiskar Nikolaus Straus, češkim jezikom zvan Mikulaš Pštros. Riječi češkog, poljskog i njemačkog jezika otisnute su frakturom.

Drugi dio Lodereckerova rječnika čini šest dvojezičnih rječnika u kojima je polazni jezik talijanski, hrvatski, češki, poljski, njemački i mađarski, a latinski je u svima drugi jezik. Taj dio ima 378 stranica i tiskao ga je također praški tiskar Ottmar, češkim jezikom zvan Jan Otmar Jakubuv Dačicky. Stranice su uređene u šest stupaca. U neparnim stupcima nalaze se riječi dotičnog jezika poredane abecednim redom, a do njih u parnome stupcu stoji prijevod na latinski. Smatra se da je ovdje prvi put tiskom objavljen dvojezični rječnik s hrvatskim kao polaznim jezikom.

3.4. Jakov Mikalja i njegov rječnik *Blago jezika slovinskoga*

3.4.1. Životopis Jakova Mikalje

Jakov Mikalja,¹⁸ odnosno na talijanskom Giacomo Micaglia te na latinskom Jacobus Micalia, hrvatski je jezikoslovac. Istraživanja su pokazala da je vjerojatno podrijetlom moliški Hrvat, a ne Talijan kako se dugo smatralo. Danas se zna da je rođen u Peschici (Pještica) u Italiji 31. ožujka 1601. godine, a

¹⁸Ravlić, Slaven, ed. "Mikalja, Jakov." Zagreb: Leksikografski zavod Miroslav Krleža, 2013. <http://www.enciklopedija.hr/Natuknica.aspx?ID=40703> (31. 8. 2014.).

umro u Loretu u Italiji 1. prosinca 1654. Prezime je dobio po očevu imenu u genitivu.¹⁹

U isusovački je novicijat stupio 1628. u Rimu. Zna se da je u Dubrovniku od 1630. do 1633. u isusovačkoj gimnaziji poučavao gramatiku. Zaredio se 1635., a do 1636. u Rimu je studirao teologiju. Od 1637. do 1645. bio je misionar u Temišvaru koji je tada bio pod osmanskom vlašću te kasnije u Trnavi u Slovačkoj. Do kraja života u Loretu je obnašao je službu ispovjednika hrv. hodočasnika.

Autor je više djela koja se tiču hrvatskoga jezika. Prvo je djelo, s naslovom *O temeljima gramatike na ilirski prilagođene (De institutione grammatica pro illyricis accommodata)*, 1637., ujedno i prva hrvatska obradba latinske gramatike E. Alvaresa, pa iako se u tom djelu autor ne spominje, predgovor, slovopis i pisma iz toga doba dokazuju Mikaljino autorstvo. Mikalja je također prvi napisao gramatiku jednoga romanskoga jezika na hrvatskome: *Gramatika talijanska u kratko ili Kratak nauk za naučiti latinski jezik* koja je ujedno i prvi udžbenik stranoga jezika na hrvatskome u nas, te su time položeni temelji hrvatskomu gramatičkom nazivlju.

Važno mu je djelo i gramatika koja je izdana i posebno kao knjižica 1649. i zajedno s hrvatsko-talijansko-lat. rječnikom *Blago jezika slovinskoga ili Solovnik u komu izgovaraju se rječi slovinske latinski i dijački* (1651.), koji predstavlja glavno Mikaljino djelo.

Analizom njegovih djela, vidi se da je Mikalja dobro poznao hrvatska književna djela i narodne govore, a posebice Bartola Kašića i njegove tekstove.

¹⁹Horvat, Vladimir. Jakov Mikalja, Isusovac–leksikograf. Zagreb, Filozofsko teološki Institut Družbe Isusove. Obnovljeni život (56) 4 (2001) str. 495-502.

Svojim je djelom utjecao na kasnije leksikografe i učvrstio izbor štokavske dijalektne osnove za budući standardni jezik. Bavio se i pravopisom, te je svoju raspravu *Od ortografije jezika slovinskoga ili načina od pisanja (De orthographia pro lingua illyrica)* objavio u svojoj knjizi *Bogoljubna razmišljanja od očenaša* (1642.) i uz *Blago jezika slovinskoga*.

3.4.2. Blago jezika slovinskoga illi Slovník u Komu izgovarajuse rjeci slovinske Latinski i Diacki. Thesaurus linguae Illyricae sive Dictionarium Illyricum. In quo verba Illyrica Italice, et Latine redduntur

Blago jezika slovinskoga illi Slovník u Komu izgovarajuse rjeci slovinske Latinski i Diacki (prema današnjem pravopisu: *Blago jezika slovinskoga ili slovník u komu se izgovaraju riječi slovinske latinski i dijački*) je trojezični rječnik, tiskan u tisuću primjeraka, s hrvatskim kao polaznim te talijanskim i latinskim kao ciljanim jezicima.²⁰ Pod pojmom „jezik slovinski“ Mikalja ne govori čiji je to jezik, već da se govori u raznim slavenskim krajevima i spominje da je od mnogo dijalekata koji se u slovinskoj zemlji govore, najljepši bosanski jezik. Ipak, analizom natuknica, vidi se da se pojam slovinski odnosi na hrvatski jezik. Termin „latinski“ znači talijanski jezik, a „dijački“ znači latinski jezik. Prema tome na današnjem hrvatskome jeziku naslov bi glasio: *Blago jezika hrvatskoga ili rječnik u kome se hrvatske riječi izgovaraju talijanski i latinski*.

²⁰Gabrić-Bagarić,

Darija.

Četiriishodištahrvatskogastandardnogajezika.

Rijeka.

OdsjekzakroatistikuFilozofskogafakultetaSveučilištauRijeci.Fluminensia 22 (2010), br. 1.str 149-162.

Naime, na talijanskom jeziku „Hrvat“ je „Croata“, a na latinskom jeziku prvo mu je značenje „Illyricus“, dakle Ilir, a drugo je značenje opet Croata, dakle Hrvat. To znači da je stvarno značenje pojma jezik slovinski prije svega jezik hrvatski, a u praktičnom smislu jezik slovinski odnosi se na sve stanovnike svih (južno)slavenskih zemalja, budući da je rječnik bio namijenjen prije svega talijanskim misionarima.

Iako je po koncepciji to opći rječnik, bilježi različito nazivlje: glazbeno, jezikoslovno, medicinsko, kulinarско, pravno, pomorsko, nazive s područja likovne umjetnosti i poezije, fitonime, i zoonime. Mjestimično *Blago jezika slovinskoga* prelazi okvire općeg rječnika prvenstveno zbog obilja enciklopedijsko-leksikonskih natuknica, te posebno bogate onomastičke građe.²¹

Rječnik, namijenjen prvenstveno pouci učenika i isusovačkoga pomlatka, donosi oko 25 000 hrvatskih riječi uglavnom u jekavskom obliku uz nešto štokavskih i čakavskih ikavizama.

3.5. **Ardelio della Bella i njegov *Dizionario italiano-latino-illirico***

3.5.1. **Životopis Ardelija della Belle**

Ardelio della Bella^{22,23} (Foggia, 1655. – Split, 1737.) bio je član Družbe Isusove, propovjednik i misionar, znanstvenik i pisac, leksikograf i

²¹Ibid.

²²Bago, Petra. “Označavanje strukture rječnika teksta prvog sveska drugog izdanja della Bellinog talijansko-latinsko-hrvatskog rječnika iz 1785.” Master thesis, Faculty of Humanities and Social Sciences, University of Zagreb, 2008.

gramatikograf. Jedan je od rijetkih stranaca u Hrvatskoj koji je svoju dužnost obavljao predano i s ljubavlju. Nives Sironić-Bonefačić u svojoj je disertaciji ukratko opisala della Bellu u njegovim ulogama:

Njegovo djelovanje utkano je u kulturnu i literarnu povijest ne samo Dubrovnika veći Splita i Zadra i ostalih krajeva kroz koje je prolazio. (...)

Svojim neumornim misionarskim radom i propovijedima na našem i talijanskom jeziku širio je među pukom i obrazovanom vlastelom ideje kršćanstva i crkvene protureformacije. Kao nastavnik pridonio je obrazovanju dubrovačke inteligencije, a kao aktivan znanstveni radnik vrlo je pažljivo pratio razvitak književnosti i jezika koji mu nije bio materinji. Svoje znanje o našem jeziku crpao je iz literature i izravnih dodira s tadašnjim stručnjacima za jezična pitanja D. Matijaševićem, I. Đurđevićem, M. Zmajevićem, itd.²⁴

Suvremenici su ga nazivali *apostolom Dalmacije* jer, pored toga što je u više od 40 godina djelovao u Dubrovniku i u cijeloj Dalmaciji, u Splitu je 1722. godine osnovao kuću misionara i bio zaslužan što su drugi misionari nastavili njegov rad sve do ukinuća Isusovačkog reda 1773.

Podrijetlom je iz stare i istaknute talijanske obitelji iz Firenze, ali se rodio u Foggiji 14. veljače 1655. godine²⁵, gdje se obitelj preselila radi posla.

²³Nives Sironić-Bonefačić napominje da postoje dvije različite grafije Ardelijevog prezimena: della Bella i Della Bella. Autorica smatra da je prihvatljiviji prvi oblik (Sironić-Bonefačić 1986 : 12); taj je oblik prihvaćen u ovom radu.

²⁴Sironić-Bonefačić, Nives. "DELLA BELLA, Ardelio." Edited by Trpimir Macan. Zagreb: Leksikografski zavod Miroslav Krleža, 1993. <http://hbl.lzmk.hr/clanak.aspx?id=4507>, (str. 33).

²⁵Nives Sironić-Bonefačić je naišla na dva podatka o datumu rođenja Ardelija della Belle: 14. veljače 1655 i 2. veljače 1655. U djelu *Isusovci i hrvatska kultura* kao datum rođenja naveden je 2. veljače 1655. godine, dok je na mrežnim stranicama Instituta za hrvatski jezik i jezikoslovlje naveden 14. veljače 1655. Budući da je jedino Sironić-Bonefačić uopće navela da je dvojben podatak o datumu rođenja Ardelija della Belle, te je u svom

Niže obrazovanje stekao je u rodnom gradu, a na studij filozofije i prava odlazi u Napulj. Godine 1677. u Rimu je stupio u Isusovački red, a 1681. ga poglavari šalju u Dubrovnik za profesora na isusovačkom kolegiju, gdje je sljedeće tri godine predavao latinski jezik i književnost te govorništvo, ali i učio hrvatski jezik. U Rim se vratio 1684. godine kako bi nastavio studij na Gregorijanskom sveučilištu. Međutim, della Bella se vidio u ulozi misionara, a ne profesora na učilištu. Stoga je na vlastiti zahtjev u Dubrovniku radio kao misionar od 1688. do 1691. godine, s kratkim prekidom 1689. godine. Ostatak života htio je posvetiti misijama u našim dalmatinskim krajevima, zbog čega se i odrekao službe profesora filozofije u Rimu. Godine 1694. otišao je u Dubrovnik gdje je djelovao kao propovjednik i misionar, a od 1696. do 1702. godine bio je rektor Dubrovačkog kolegija. Osim toga, dovršio je zgradu kolegija i započeo gradnju isusovačke crkve sv. Ignacija. Nakratko se, sredinom 1702. godine, na godinu dana vraća u Italiju kako bi predavao skolastičku teologiju u Firenzi. Konačno se krajem 1703. godine vraća za misionara u Split. Narednih 30 godina della Bella je posjetio 17 biskupija duž cijele Dalmacije, od Cresa do Kotora, a djelovao je i u Hercegovini. Preminuo je 3. prosinca 1737. godine u Splitu od posljedica moždanog udara koji je doživio četiri godine ranije.

Kao što je već spomenuto, della Bella nije bio samo propovjednik i misionar već i znanstvenik i pisac. Dvadesetak godina tijekom boravka u Hrvatskoj prikupljao je građu za svoj rječnik *Dizionario italiano-latino-illirico*, koji je objavio u Veneciji 1728. godine. Petar Bašić iz Dubrovnika poboljšao je i priredio za tisak drugo izdanje della Bellinog rječnika, koje je objavljeno u dva sveska u Dubrovniku 1785. godine. Osim toga, u Veneciji je 1805. godine Matija Čulić posthumno objavio njegove propovjedi na hrvatskom jeziku

djelu ukratko argumentirala zašto smatra da je točan prvi, a ne drugi datum (Sironić-Bonefačić 1986 : 12), odlučeno je da će se usvojiti njezin podatak.

održane u bratovštini Dobre smrti u Dubrovniku, pod nazivom *Razgovori i propovidagna oza Ardelia della Belle*.

3.5.2. Dizionario italiano-latino-illirico A cui si premettono alcune brevi Istruzioni Gramaticali, necessarie per apprendere la Lingua e l'Ortografia Illirica

Svrhu ovog rječnika možemo naći u ideji vodilji isusovačkog reda. Njihova glavna ideja vodilja *se temelji na saznanju da se vjera u narodu može širiti isključivo na narodnom jeziku. Zato su isusovački misionari morali naučiti jezik naroda među kojim djeluju*²⁶. Ardelio della Bella je za vrijeme prvog boravka u Dubrovniku primijetio da nedostaje priručnik koji bi talijanskim misionarima pomogao pri širenju vjere na narodnom, odnosno hrvatskom jeziku. Zato je odlučio sastaviti rječnik *Dizionario italiano-latino-illirico*, te ga *namijenio prvenstveno budućim misionarima u slavenskim zemljama, da bi uz rječnik i gramatiku što potpunije i brže naučili naš jezik*²⁷. Prvo izdanje rječnika izašlo je u Veneciji 1728. godine. U tom se izdanju, uz talijansko-latinsko-hrvatski rječnik, nalazi i latinsko-talijanski indeks. U drugom izdanju rječnika, koji je poboljšao i priredio za tisak Petar Bašić, izostavljen je latinsko-talijanski indeks. Drugo izdanje objavljeno je u Dubrovniku 1785. godine u dva sveska. U ovom radu upotrijebljeno je upravo drugo izdanje della Bellinog rječnika.

Rječnik spada u skupinu višejezičnih rječnika. Sastavljen je od sveukupno 899 stranica. Sastoji se od dva dijela. Prvi dio je uvod koji je pisan je

²⁶Sironić-Bonefačić, Nives. "DELLA BELLA, Ardelio." Edited by Trpimir Macan. Zagreb: Leksikografski zavod Miroslav Krleža, 1993. <http://hbl.lzmk.hr/clanak.aspx?id=4507>, (str. 36).

²⁷Sironić-Bonefačić, Nives. "DELLA BELLA, Ardelio." Edited by Trpimir Macan. Zagreb: Leksikografski zavod Miroslav Krleža, 1993. <http://hbl.lzmk.hr/clanak.aspx?id=4507>, (str. 36).

na talijanskom jeziku i nalazi se na prvih 56 stranica paginiranih rimskim brojevima. Drugi dio je popis riječi na 843 stranice paginirane arapskim brojevima. Rječnik se može također podijeliti na dva dijela prema svescima. U tom slučaju prvi svezak sadrži 451 stranicu u koju spada uvodni dio i popis riječi od slova „A“ do slova „H“. Drugi svezak ima 448 stranica na kojima je popis riječi od slova „I“ do slova „Z“.

Na prvom dijelu uvoda, na *iii.* i *iv.* stranici nalazi se posveta *Dobročiniteljima (ECCELLENZE)*, koju potpisuje Carlo Antonio Occhi. Occhi je bio Mlečanin koji je osnovao prvu tiskaru u Dubrovniku 1783. godine. Posveta je namijenjena dubrovačkom Senatu²⁸ u kojoj Occhi naglašava značenje dubrovačkoga govora u razvoju hrvatskog jezika, uspoređujući ga sa značenjem atičkog dijalekta u starih Grka²⁹.

Zatim slijedi predgovor urednika (*PREFAZIONE DELL' EDITORE*) od *v.* do *x.* stranice u kojem Petar Bašić govori o tome kako je priredio drugo izdanje della Bellinog trojezičnog rječnika i ispravio pogreške. Navodi kako je uveo tri novosti:

- dodao je nove riječi
- promijenio je ortografiju
- izbacio nepotrebne stvari i objasnio zašto je izbacio neke riječi.

²⁸Zanimljivost o dubrovačkom Senatu: Naime, Senat je nakon tiskanja prvog sveska Della Bellinog rječnika, koji je zbog posvete morao novčano potpomoći, početkom 1786. godine donio odluku da mu ubuduće nitko neće smjeti posvećivati knjige ukoliko za to nije dobio dopuštenje. (Čučić, V. Prvi tiskari u Dubrovniku s popisom tiskane grade. // Vjesnik bibliotekara Hrvatske. 48, 3/4(2005), str. 108-158.), (str. 15).

²⁹Čučić, V. Prvi tiskari u Dubrovniku s popisom tiskane grade. // Vjesnik bibliotekara Hrvatske. 48, 3/4(2005), str. 108-158, (str. 119).

Gramatički pregled *ilirskog* jezika je ostavio, kao što je bio i u prvom izdanju. Slijedi kratko tumačenje grafije; nije slijedio della Bellinu grafiju, već grafiju opata Ignjata Đurđevića (*Ignazio Gjorgi*). Nakon toga priča o tome što je sve nadodano i zašto je ovo izdanje izašlo u dva toma. Na kraju govori o citatima, te o našem jeziku, dubrovačkom i bosanskom govoru. Della Bella je u prvom izdanju rječnika iznosio ideje *o ilirskom kao općeslavenskom jeziku koji se govori u velikom dijelu Evrope i Azije*³⁰, te je tumačio zašto je odabrao štokavsko narječje kao osnovu hrvatskog prijevoda, točnije njegove najraširenije jekavske varijante dubrovačkog i bosanskog govora.³¹ Jezik rječnika je štokavska jekavica, ali della Bella nije izostavio ni ostala dva hrvatska narječja.

Zatim od *xi.* do *xiii.* stranice dolazi kazalo autora i djela citiranih u rječniku, te njihove kratice korištene uz navedene citate (*Indice Degli Autori Antichi, e Moderni citati in questo Vocabolario colla spiegazione delle Abbreviature, onde sono indicati; e molte voci de' quali Autori si trovano anche senza le Citazioni*), koje su znale varirati zato što je rječnik nastajao tijekom dugog razdoblja. Della Bella je po prvi put u hrvatskoj leksikografiji uz natuknice unio i potvrde, odnosno primjere iz književnih djela i iz usmene narodne književnosti u kojima se ta riječ pojavljuje. Može se naći oko 40 djela štokavskih i čakavskih pisaca 16. i 17. stoljeća. Povrh toga je uvrstio narodne poslovice i fraze. Pored poznatijih pisaca starije hrvatske književnosti, primjerice Ivana Gundulića, Marina Držića ili Petra Hektorovića, i njihovih značajnih djela (*Osman, Tirena* ili *Ribanje i ribarsko prigovaranje*), pojavljuju

³⁰Nives Sironić-Bonefačić u svojoj doktorskoj disertaciji podrobnije objašnjava da ideje o rasprostranjenosti ilirskog i njegove upotrebe kao panslavenskog jezika možemo naći u mnogim djelima nastalim prije della Bellinog rječnika, ali i u djelima della Bellinih suvremenika.

³¹Sironić-Bonefačić, Nives. "DELLA BELLA, Ardelio." Edited by Trpimir Macan. Zagreb: Leksikografski zavod Miroslav Krleža, 1993. <http://hbl.lzmk.hr/clanak.aspx?id=4507>, (str. 56), (31. 8. 2014.).

se i danas zamalo zaboravljeni pisci kao što su Dubrovčanin Marin Burešić te Splitsanin Aleksandar Komulović. Zanimljiv je podatak da je Gundulićev Osman tiskan gotovo stotinu godina nakon rječnika, pa su citati iz njega ispisani iz rukopisnih predložaka. Njihovo navođenje u rječniku prvo je objavljivanje njihovog teksta.³² Popis autora i djela korištenih kao korpus za izradu rječnika nalazi se u prilogu.

Na stranici *xiv* slijedi tablica svih ostalih kratica korištenih za opis gramatičkih vrsta u rječniku i gramatičkim uputama (*TAVOLA D' alcune altre abbreviature, che si trovano in quest' Opera*). Popis prethodno navedenih kratica nalazi se u prilogu.

Potom od *xv.* do *xvii.* stranice slijede upute o pisanju i čitanju hrvatskih riječi (*AVVERTIMENTI Per leggere con facilità le parole Illiriche scritte con Lettere Italiane*). S obzirom na to da je rječnik namijenjen talijanskim misionarima, jasno je zašto su hrvatske riječi pisane talijanskim pravopisom. Urednik Petar Bašić govori da se u rječniku mogu naći dva načina pisanja hrvatskih riječi talijanskom grafijom: prvi je način Ardelija della Belle, a drugi Ignjata Đurđevića. Na temelju tih dvaju načina nastale su upute za lakše čitanje, u kojima urednik navodi kako bi određenu riječ napisao della Bella, a kako Đurđević, te govori što je odabrao za ovo izdanje.

Na kraju uvoda, od *xviii.* do *lvi.* stranice, nalazi se pregled gramatičkih pravila hrvatskog jezika (*ISTRUZIONI GRAMATICALI DELLA LINGUA ILLIRICA*). Pri sklapanju rječnika della Bella se ugledao na jednog drugog isusovca, Jakova Mikalju (Italija, 17. st.) koji je objavio trojezični hrvatsko-talijansko-latinski rječnik pod nazivom *Blago jezika slovinskoga* (Loreto-

³²Ibid. (str. 64).

Ancona, 1649. – 1651.). Naime, Mikalja je na početak rječnika stavio kratku gramatiku hrvatskog jezika pisanu prema gramatici Bartola Kašića.³³ Della Bella je također na početak rječnika uvrstio kratku uvodnu gramatiku *ilirskog* jezika, koju je urednik drugog izdanja i zadržao. Gramatika je kao samostalno djelo pod nazivom *Principi elementari della grammatica illirica* objavljena u Dubrovniku 1837. godine.

Konačno slijedi trojezični rječnik. Popis riječi od slova „A“ do slova „H“ nalazi se u prvom svesku od 1. do 395. stranice. Riječi od slova „I“ do slova „Z“ nalaze se u drugom svesku od 3. do 448. stranice. Stranica je podijeljena na dva stupca. Natuknica počinje s riječju na talijanskom jeziku, nakon koje slijedi prijevod na latinski, a zatim na hrvatski jezik. Riječi na talijanskom pisane su uspravnim slovima. Riječi na latinskom pisane su kurzivom. Riječi na hrvatskom pisane su ponovo uspravnim slovima, dok su primjeri iz književnih djela i iz usmene narodne književnosti uvučeni i pisani kurzivom. Natuknice su posložene abecednim redom. U prvom ih svesku ima točno 7 973, a u drugom 10 999 (budući drugi svezak rječnika nije još do kraja strukturiran, može se dogoditi da ta brojka nije sasvim točna, ali je za to vjerojatnost mala). U oba dijela rječnika postoje, dakle, 18 972 natuknice što della Bellin rječnik čini srednje velikim rječnikom, a za svoje doba velikim.

Budući da su na nekim riječima stavljeni i akcenti, rječnik može poslužiti i kao ishodište za istraživanje povijesti akcentuacije u hrvatskom jeziku.

³³Bartol Kašić (Hrvatska, Italija, 16./17. st.) objavio je prvu hrvatsku gramatiku *Institutionum linguae illyricae – libri duo* 1604. godine u Rimu po naređenju tadašnjeg generala Družbe Isusove. Naime, svrha gramatike je bila pomoći Rimskome kolegiju pri obrazovanju mladih isusovaca koji su nakon školovanja svoju misiju obavljali na Balkanu; zato je gramatika pisana na latinskom jeziku (Korade et. al. 1993:49).

3.6. Joakim Stulli i njegovo *Rjecsosloxje ilirsko (slovinsko)-italiansko-latinsko*

3.6.1. Životopis Joakima Stullija

Dubrovački je leksikograf franjevac Joakim Stulli (Stulić)³⁴ rođen u Dubrovniku 1730. godine, gdje je i umro 1817. godine. Poslije osnovnog školovanja u rodnome gradu pohađao je isusovački kolegij, a zatim nastavio filozofsko-teološki studij u samostanu Male braće. Zatim je tri godine studirao bogoslovlje u Rimu na franjevačkom studiju u samostanu Ara Coeli nakon čega se vratio u Dubrovnik gdje je od 1760. radio na svojem velikom rječniku koji je objavljen u šest svezaka u tri knjige: *Lexicon Latino-Italico-Illyricum*. U radu se ugledao na rječnik Ardelija della Belle, na talijanski *torinski rječnik* te na druge rječnike.³⁵ Kao građu koristio se primjerima koje je ispisivao iz književnih djela dubrovačkih, dalmatinskih, hercegovačkih, bosanskih, slavonskih i istarskih pisaca, ali i iz drugih izvora (na primjer iz rječnika Ardelija della Belle, Ivana Belostenca i Andrije Jambrešića), pa je tako i pojam „ilirski“ Stulli shvaćao i šire nego što je nama danas hrvatski. Rječosložje je najveće djelo starije hrvatske leksikografije, te je jako je utjecalo na preporodno jezično djelovanje.

³⁴Prema članku iz Wikipedije „Joakim Stulić“, http://en.wikipedia.org/wiki/Joakim_Stuli%C4%87 (3. 9. 2014.).

³⁵Prema članku iz Proleksis enciklopedije „Stull (Stulić) Joakim“, <http://proleksis.lzmk.hr/47559/> (3. 9. 2014.).

3.6.2. Rjecsosloxje ilirsko (slovinsko)-italiansko-latinsko u komu donosuse upotrebljenja, urednia, mucsnia istieh jezika krasnoslovja nacsini, izgovaranja i prorjecsja

Rječnik Joakima Stullija, pod naslovom *Rjecsosloxje ilirsko (slovinsko)-italiansko-latinsko u komu donosuse upotrebljenja, urednia, mucsnia istieh jezika krasnoslovja nacsini, izgovaranja i prorjecsja*, najveći je rječnik u povijesti starije hrvatske leksikografije (ukupno 4 721 stranica). Njegov rječnik, na kojemu je radio oko pola stoljeća, ima tri dijela, a svaki dio čine dva sveska. Dijelovi su izašli pod sljedećim naslovima: *Lexicon latino-italico-illyricum* (tiskan u Budimu 1801. godine), *Rječosložje iliričko-italijansko-latinsko* (tiskan u Dubrovniku 1806. godine), *Vocabolario italiano-illyrico-latino* (tiskan u Dubrovniku 1810. godine).

Polazni jezik prvoga dijela je latinski, drugoga hrvatski, a trećega talijanski.

Građu za rječnik Stulli je uzimao iz svih postojećih hrvatskih rječnika te iz književnih djela, djela usmene narodne književnosti, kao i iz tiskanih i rukopisnih djela različite namjene i stilova. Zanimljivo je da kao izvor nije uzeo Vrančićev rječnik, a također ni djela Marka Marulića, Hanibala Lucića, Petra Zoranića, Petra Zrinskog, Frana Krste Frankopana ili Pavla Rittera Vitezovića.

Temeljni leksik rječnika je štokavski leksik, međutim rječnik sadrži i leksik čakavskoga i kajkavskoga narječja te staroslavenizme, rusizme, bohemizme, polonizme, riječi slovenskoga jezika i osobne autorove kovanice

budući da je smatrao da je ilirski jezik slavenski jezik koji razumiju svi Slaveni pa ga je želio učiniti bogatijim.³⁶

3.7. Ivan Mažuranić i Jakov Užarević te njihov *Njemačko-ilirski slovar*

3.7.1. Životopis Ivana Mažuranića i Jakova Užarevića

Ivan Mažuranić, hrvatski pjesnik i političar, rođen je 11. kolovoza 1814. godine u Novom Vinodolskom, a umro 4. kolovoza 1890. u Zagrebu.³⁷ Nakon završenog Pravnog fakulteta zaposlio se kao nastavnik u gimnaziji u Zagrebu, a kasnije postaje odvjetnik u Karlovcu. 1850. godine imenovan je državnim odvjetnikom.

Usporedo razvija i svoju političku karijeru koju je započeo kao gradski zastupnik u Karlovcu, a završava je 1880. godine u funkciji bana pučanina. Jedan od značajnijih jezično-političkih događaja za vrijeme njegove političke karijere je tzv. Bečki dogovor, ili sporazum pet hrvatskih pisaca i filologa (Ivan Mažuranić, Dimitrije Demeter, Vinko Pacel, Stjepan Pejaković, Ivan Kukuljević Sakcinski) i dva srpska (Vuk Karadžić, Đuro Daničić), u organizaciji poznatog slovenskog slavista Franca Miklošiča.

Taj je sporazum nastao kao posljedica inicijativa austrijske vlade da se ujednači pravna terminologija, jer je „zemaljski jezik“ na više područja prestajao biti njemački. U biti, radilo se o shvaćanjima da su hrvatski i srpski jedan jezik,

³⁶Ibid.

³⁷Prema članku iz Wikipedije „Ivan Mažuranić“, http://en.wikipedia.org/wiki/Ivan_Ma%C5%BEurani%C4%87 (3. 9. 2014.).

pa ga treba ujednačiti u svim vidovima. Sporazum je često kasnije apostrofirao kao povijesna prekretnica, no realno nije imao većega utjecaja u praksi jer su se hrvatski pisci i filolozi i dalje držali ilirskoga oblika jezika, a srpski su uglavnom napustili ili marginalizirali ijekavicu.

Pjesme je počeo pisati kao srednjoškolac u Rijeci. U razdoblju od 1835. do 1841. godine surađivao je s Gajevom Danicom, a pet godina kasnije objavio je svoje najznačajnije djelo, klasični junački ep „Smrt Smail-age Čengića“. Nakon prekida suradnje s Danicom izdao je „Němačko-Ilirski slovar“ koji je nastao u suradnji s dr. Jakovom Užarevićem.

O suautoru ovoga rječnika **Jakovu Užareviću** (1808. – 1881.) ima vrlo malo podataka.³⁸ Bio je doktor znanosti, jedan od začetnika narodnog buđenja pod vodstvom Ljudevita Gaja, tajnik Matice ilirske od 1842. do 1845. godine, urednik Narodnih novina od 1850. do 1859. godine te urednik časopisa „Kolo“. Zna se da je studirao u Beču te je bio prijatelj s Dimitrijom Demetrom. Prevodio je i s njemačkog, pa se tako zna za njegov prijevod s njemačkog *Kratak opis Cesarevine austrijske u obće, i pojedinih nejzinih pokrajinah na pose : za niže gimnazije i niže realke* koji je sastavio Ludovik vitez Heufler, a izdan je u Beču 1861. godine.³⁹

³⁸Na službenim stranicama općine Petrijevci, govori se o Jakovu Užareviću kao istaknutom Petrijevčaninu. <http://www.petrijevci.hr/zanimljivosti.php?id=5> (3. 9. 2014.).

³⁹Katalog HAZU. Iz starog kataloga: *Kratak opis Cesarevine austrijske u obće, i pojedinih nejzinih pokrajinah na pose : za niže gimnazije i niže realke* / sastavio Ludovik vitez Heufler ; prev. Jakov Užarević. - U Beču : U C. kr. nakl. školskih knjigah, 1861. - XIII, 389 str. : ilustr., bakrorez s grbom cara Franje Josipa I. ; 8° (22 cm). <http://katalog.hazu.hr/WebCGI.exe?Tip=Listic&Jbmg=097981&Baza=1> (3. 9. 2014.).

3.7.2. Deutsch-ilirisches Wörterbuch. Njemačko-ilirski slovar

Kao što je već navedeno, ovaj je njemačko-hrvatski rječnik pod naslovom *Deutsch-ilirisches Wörterbuch. Njemačko-ilirski slovar* nastao u suautorstvu Ivana Mažuranića i Jakova Užarevića, te izdan u Zagrebu 1842. godine,⁴⁰ specifičan po tome što su njemački izrazi pisani goticom. Sam rječnik ima 486 stranica i sastoji se od nekoliko dijelova:

Predgovor govori o razlici između duha njemačkog i hrvatskoga jezika te objašnjava zašto su neki njemački izrazi ponekad opisno prevedeni na hrvatski jezik. Pravilo za kojim su se autori povodili prilikom stvaranja ovog rječnika je poštovanje duha obaju jezika. Ovdje donosimo cjeloviti tekst hrvatskoga predgovora:

P R E D G O V O R .

Duh nêmačkoga jezika različit je jako od duha ilirskoga jezika. Što Nêmac kadkada kaže jednom rěči, to mi nemoremo neg s više; a što mi često jednom rěči izgovorimo, tomu trëba u njega kad – god i tri i četiri. A to se još većma ukazuje u frazah, u posloviceh itd. S toga nastojasmo, da i u tom razliku izmedju ova dva jezika pokažemo, u koliko nam biaše moguće. U obće nastojali smo pre – vodit ne samo gole rěči, nego i duh od jezika.

U nas je osobito velika nevolja, kad imaš prevesti koi posebni izraz; jer se u nas dosele niti u školi, niti u dëlarnici nebesëdjaše jezikom domaćim ništa, ili jako malo. To polje je još sasvim neob – radjeno. Zato će po svoj prilici bit u ovoj knjizi najviše pogrë – šakah, što se tiče izrazah poslovnih, a osobito travarstva; jer kako sve znanosti, tako ti je obradjena u nas i botanika, to će reć ni – kako. Što imamo dobra, to spada svekoliko u krug od običnoga svakdanjega jezika. I to je osobito ono, što smo nastojali najve – ćom pomnjom obraditi; jer je to temelj svakoga jezika; to je ko – rën i stablo, koje kad ojača jedared, pustit će lasno i granje na sve strane i prolistati. Kad naši budu razumëli dobro svoj svak – danji jezik, ako budu dobro i temeljito vëšti kojoj nauci i od na –

⁴⁰Mažuranić, Ivan i Jakov Užarević. **Deutsch-ilirisches Wörterbuch. Njemačko-ilirski slovar**, Verlag und Druck der k. priv. Ilirischen National-Buchdruckerei von Dr. Ljudewit Gaj. Zagreb, 1842.

ravi k tomu pozvani, napraviti će oni lasno svekoličke izraze, koji su nam još potrebni. Što ne učini jedan, hoće drugi. Samo valja da bude čvrsto temelj.

Što se u ovoj knjizi govori o njemačkoj slovnici, uzeto je iz Slovara njemačko–francuskoga od Henšela; njemu dakle i ukor i slava; jer mi njemački slovničari nit jesmo, nit hoćemo da budemo.

I. Mažuranić

J. Užarević

Gramatika njemačkog jezika – U predgovoru se nalazi napomena da je njemačka gramatika (njemačka slovnica) preuzeta iz Henšelovog Njemačko-francuskog rječnika. Sama gramatika obrađuje sljedeće cjeline: abecedu, članove (muški, ženski, srednji – u rječniku nazvan nikakov) i njihovu deklinaciju, deklinaciju imenica svih triju rodova u jednini i množini (imenice s određenim, neodređenim članom ili bez člana), komparaciju pridjeva, brojeve, zamjenice, glagole i njihovu konjugaciju.

Popis kratica – u rječniku stoje iza riječi na njemačkom jeziku i određuju vrstu riječi. U slučaju imenica određuju njihov rod (muški, ženski, srednji), a kod glagola određuju radi li se o aktivnom, aktivnom i neprijelaznom glagolu, te prijelaznom ili povratnom glagolskom obliku. U slučaju ostalih pojmova ne postoji daljnja klasifikacija osim određivanja vrste riječi.

Glavni dio rječnika – pojmovi su abecednim redoslijedom poredani u dva stupca. Uz njemačku riječ pisanu goticom nalazi se kratica koja označava vrstu riječi, pisana kurzivom, nakon čega slijedi prijevod na hrvatski (ilirski jezik) pisan latiničkim slovima. Primjer:

Aalpuppe, f. vabilo za lovit jegulje.

odnosno, pisano izvornom grafijom

Aalpuppe, f. vabilo za lovit jegulje.

Uz prijevod životinjskih i biljnih vrsta naveden je i njihov latinski naziv.

Primjer:

Aalfrau, f. jeguljarica; (ein Fiſch) balavica
(riba). *Blennius viviparus* Linn.

odnosno, pisano izvornom grafijom:

Aalfrau, f. jeguljarica; (ein Fiſc) balavica
(riba). *Blennius viviparus* Linn.

Popis zemljopisnih naziva – pisan je u istom obliku kao i ostale riječi.

Konjugacija nepravilnih glagola – tabelarni prikaz nepravilnih glagola u sljedećim oblicima: 1. lice jednine aorista, particip perfekta koji se koristi u tvorbi perfekta i oblik glagola u infinitivu. Primjer:

Aß, Gegeſſen, Eſſen.

3.8. Mirko Divković i njegov *Latinsko-hrvatski rječnik za škole*

3.8.1. Životopis Mirka Divkovića

Mirko Divković,⁴¹ gramatičar, leksikograf i prevoditelj, rođen je 1843. u Zagrebu, a umro je 1924. u Samoboru. Gimnaziju je završio 1863. u Zagrebu, a studij klasične i slavenske filologije 1867. u Beču. Bio je nastavnik u

⁴¹Sastavljen prema članku Divković Mirko (autor: Branko Plešić i Redakcija) u: Hrvatski biografski leksikon (ur. Trpimir Macan), Online izdanje, Leksikografski zavod Miroslav Krleža, Zagreb, 1983-2009. <http://hbl.lzmk.hr/clanak.aspx?id=4861> (3. 9. 2014.).

varaždinskoj gimnaziji te, napokon, promijenivši više zaposlenja, postao ravnateljem Klasične gimnazije u Zagrebu. Od 1908. živio je u Samoboru.

Proučavao je hrvatski i latinski jezik te je napisao više školskih priručnika, među kojima posebno mjesto zauzima više puta prerađivana i dopunjavana Hrvatska gramatika. Njen prvi dio, *Oblici*, pojavio se 1879., a drugi dio, *Sintaksa*, 1881; od 1898. oba djela tiskaju se zajedno pod nazivom *Oblici i sintaksa hrvatskoga jezika*. Rađena je prema Daničićevim gramatičkim djelima i Miklošičevoj poredbenoj slavenskoj sintaksi, a razdio o prijedlozima napisao je August Musić.

Godine 1877. izdao je *Rječnik hrvatsko-latinski*. Kao suradnik prvog izdanja Latinsko-hrvatskog rječnika Sebastijana Žepića iz 1881., obradio je slova I-N i T-Z, a 1900. pripravio je drugo dopunjeno i prošireno izdanje, djelom akcentuirano, obogativši ga frazeologijom i sinonimijom. Djelo je pretisnuto 1980., a njegov priređivač Veljko Gortan ističe vrsne Divkovićeve etimologije i temeljito poznavanje hrvatskoga jezičnog blaga.

Prevodio je s ruskoga, ponajviše Turgenjevljevu i Tolstojevu prozu, a od 1923. objavio je u izdanju Matice hrvatske *Ruske narodne pripovijetke s ilustracijama* Ljube Babića. Kao Matičin tajnik (od 1871.) objelodanio je njezinu povijest od osnutka 1842. do 1874. (*Izvyješće o Kraljevskoj velikoj gimnaziji u Zagrebu, 1876.-77.*). Znanstvene i stručne rasprave objavljivao je u Kršćanskoj školi (1900.), *Nastavnom vjesniku* (1902., 1910., 1912. – 1213., 1914. – 1915.) i dr.

3.8.2. Latinsko-hrvatski rječnik za škole

Rječnik Mirka Divkovića⁴² pod naslovom *Latinsko-hrvatski rječnik za škole*, koji obuhvaća 1 161 stranicu, vrijedan je i dosta opsežan priručnik. U Predgovoru Rječnika srednjovjekovnog latiniteta Jugoslavije (*Lexicon latinitatis medii aevi Iugoslaviae*), što ga je izdala Jugoslavenska akademija znanosti i umjetnosti, I. sv. 1973., II. sv. 1978., istaknuto je da su u nj uvrštene i one riječi iz klasičnog latiniteta koje ne obuhvaća „naš standardni Latinsko-hrvatski rječnik za škole Mirka Divkovića, II. izdanje 1900.“ Pozivanje na taj Rječnik jasno izražava uvjerenje Redakcije tog Akademijina Rječnika da je kod nas Divkovićev Rječnik najbolji rječnik za klasični latinitet.

Vrijednost Divkovićeve Rječnika nije samo u njegovoj opsežnosti nego i u obradi. Kao stručnjak kroatist pazio je na točnost i izražajnost riječi kojima je prevodio latinske riječi i fraze. U Predgovoru napominje da je na mnogo mjesta naše riječi snabdio naglaskom „da bi se učenici, kojima nije štokavština materinji jezik, privikli pravilnom izgovoru“. Latinske riječi imaju oznake za duljinu (ˉ) i kraćinu (˘) vokala. Pri tiskanju su se tu i tamo potkrale pogreške u toj oznaci kvantitete, no to ne može utjecati na vrijednost Rječnika kao cjeline. Iako on nije zamišljen kao etimologijski, na mnogo je mjesta objašnjen postanak latinskih riječi i njihova međusobna povezanost, npr. *ampŭlla* - deminutiv od *amphora*, *aurifer* (*aurum* i *fero*), *auceps*, *cŭpis* (*avis* i *capio*), *difficilis* (*dis* i *facilis*), *nĕgōtium* (*nec* i *otium*), *pĕcūnia* - *pecus*, *sŭcŭla*- deminutiv od *sus* itd. Kod brojnih složenica uz oblik s asimilacijom konsonanata naveden je i oblik bez asimilacije, npr. *irrīdeo* (*in-rid*), *assum* (*ad-sum*), *sŭppōno* (*sub-p.*). Više

⁴²Sastavljen prema: **Gortan, Veljko.** (O latinsko-hrvatskom rječniku Mirka Divkovića). Predgovor trećem reprint izdanju latinsko-hrvatskog Rječnika za škole Mirka Divkovića iz 1900. Reprint: Naprijed, Zagreb, 1980. str. IV-V.

puta dano je i znanstveno tumačenje postanka i značenja pojedinih riječi, npr. *bīmus*, *adj.* skraćeno od *bihimus*; *bis* i *hiems* – „dvozimni“, dvogodišnji. Za riječi grčkog podrijetla ili preuzete iz grčkog jezika redovno je u zagradi navedena izvorna grčka riječ.

Povijesna, mitološka i zemljopisna imena uvrštena su u Rječnik prema abecednom redu, a nisu, kao što je bilo uobičajeno, sakupljena u posebnom odjeljku. Divkovićev je postupak u tome jednostavniji i praktičniji. Svako vlastito ime popraćeno je najhitnijim sažetim podacima, a kod zemljopisnih imena, gdje je bilo moguće, dodan je i današnji naziv.

Kako je Divkovićev Rječnik objavljen prije više od sto godina, prirodno je da se u njemu nađe po koji jezični arhaizam, npr. Mačedonija i Mačedonci, pazljiv, jastrijeb i sl., ali to ne mora smetati današnjeg čitatelja.

Najveća je vrijednost tog Rječnika u bogatoj frazeologiji i u obilju tzv. sinonima. U tome se njegov autor pokazao temeljitim poznavateljem našeg jezičnog blaga. Usp. *aestūo*, *l.* pod 2) talasati se, bibati se, burkati se, pjeniti se; *flexibilis*, *adj.* [*flecto*] — gibak, vitak, prutak, mek; *ōrnātus*, *adj.* ima ova značenja, razvrstana u četiri skupine: 1) sasvim oružan, opremljen; zauzdan; 2) kićen, lijep; 3) slavan, dostojan, ugledan, odlikovan; 4) veoma ugledan, viđen, izvrstan. Uz osnovno značenje redovno je posebno označeno i značenje u prenesenom smislu ako ga ima.

Kod nekih riječi ili fraza kraticom su navedena imena latinskih pisaca, kod kojih se javljaju, a tzv. „*Hápakṣ eireména*“ (grčki), tj. riječi koje su upotrijebljene samo jedanput, popraćene su imenom pisca i točnim citatom djela u kojem je riječ zabilježena, npr. *haesitantia*, *ae, f.* [*haesito*] — neprèstano

zàpinjânje u govòrênju, mucanje, *linguae*, jednom *Cic. Phil, 3, 6, 16*; *fictrix*, *īcis*, f. [*fictor*]graditeljica, tvoriteljica, jedn. *Cic. n. d. 3, 39, 92*.

Po uzoru na druge latinske rječnike Divković zvjezdicom označuje latinske riječi pjesničke prirode koje se u pravilu ne upotrebljavaju u proznim tekstovima, a znakom † riječi koje ne pripadaju tzv. klasičnom, nego pretklasičnom ili poklasičnom periodu latinskog jezika. Kraticom *archaist.* ili *arhaist.* obilježene su zastarjele riječi, npr. *pàvio*, 4. (*arhaist.*) nabijati, pobijati, *terram*. **Tumačenje znakova i kratica** na tri stranice daje upute za njihovo razumijevanje. Tu su ujedno navedeni i latinski pisci na temelju kojih je sastavljen Divkovićev Rječnik. To su: Aurelije Viktor, Cezar, Ciceron, Eutropije, Fedro, Horacije, Justin, Katul, Kvint Kurcije, Kvintilijan, Livije, Nepot, Ovidije, Plaut, Plinije Mlađi, Propercije, Salustije, Svetonije, Tacit, Terencije, Tibul i Vergilije.

3.9. Dragutin Antun Parčić i njegov Rječnik hrvatsko-talijanski.

3.9.1. Životopis Dragutina Antuna Parčića

Dragutin Antun Parčić^{43,44,45} (Vrbnik, 1832. – Rim, 1902.) hrvatski je jezikoslovac, filolog, glagoljaš i leksikograf. Bio je na tragu zagrebačke

⁴³Parčić, Dragutin Antun. *Hrvatska Enciklopedija, Mrežno izdanje*. Ravlić, Slaven (ur.). Zagreb : Leksikografski zavod Miroslav Krleža, <http://www.enciklopedija.hr/Natuknica.aspx?ID=40703> (29.8.2014.).

⁴⁴Gostl, Igor. *Dragutin Antun Parčić*. Mala Knjižnica Matice Hrvatske, kolo 8, knj. 47. Zagreb: Matica hrvatska, 1998.

⁴⁵Vince, Zlatko. "Leksikograf Dragutin Parčić U Svojem Vremenu." In *Rječnik Hrvatsko-Talijanski*, by Dragutin Antun Parčić, 1241-76. Zagreb: Artresor studio, 1995.

filološke škole te se protivio hrvatskim vukovcima. Za tisak je 1893. godine pripremio posljednje izdanje *Misala* tiskano glagoljskim slovima. Njegovi dvojezičnici izlaze u nekoliko dopunjenih i proširenih izdanja: talijansko-hrvatski izlazi u četiri izdanja⁴⁶, dok hrvatsko-talijanski izlazi u tri izdanja i jednim pretiskom⁴⁷. Usto, gajio je zanimanje za razna znanstvena i stručna područja. Tako se bavio tiskarstvom, tipografijom, botanikom, zemljopisom, kartografijom, astronomijom, prevođenjem, mehanikom, slikarstvom, bakrorezom i pjesništvom. Osim toga, okušao se u svim područjima fotografije te se smatra pionirom hrvatske fotografije.

Antun Parčić rođen je u Vrbniku na Krku 26. svibnja 1832. godine, u doba *ilirskog preporoda*. Pučku školu pohađao je u Vrbniku. Kao devetogodišnjak pristupio je Provinciji franjevac trećoredaca u Glavotoku i primio ime Dragutin. Za vrijeme školovanja u Zadru boravio je u samostanu sv. Mihovila, snažnom glagoljaškom središtu. Tu je razvio zanimanje za botaniku te posebno leksikografiju. Svoj prvi rječnik od 120 stranica sastavio je u 6. razredu gimnazije. Prvu tiskaru na Krku *Serafinski tisak* pokrenuo je 1875. godine u samostanu sv. Marije u Glavotoku. Radio je kao profesor, ali je nastavu i dušebrižništvo smatrao teretom, jer se htio posvetiti leksikografiji i

⁴⁶Prvo izdanje talijansko-hrvatskog rječnika tiskano je u Zadru 1868. godine pod naslovom Rječnik talijansko-slovinski (hrvatski) u nakladi braće Battara. Drugo izdanje ovog rječnika tiska nakladnik Luster u Senju 1887. godine. Treće izdanje je tiskano također u Senju u nakladi Hreljanovića 1908. godine, dok četvrto izdanje tiska nakladnik Schonfeld u Zadru 1909. godine.

⁴⁷Rječnik: Rječnik hrvatsko-talijanski tiskan u Zadru 1901. kao treće popravljeno i povećano izdanje hrvatsko-talijanskog rječnika nakladom Narodnog lista. Prvo izdanje pod naslovom Rječnik ilirsko-talijanski tiskano je 1858. godine u Zadru u nakladi Petra Abelića, dok je drugo izdanje pod naslovom Rječnik slovinsko-talijanski tiskan u Zadru 1874. godine u nakladi braće Battara. Pretisak trećeg izdanja dvojezičnika izdao je 1995. godine. Već iz naslova izdanja rječnika moguće je pratiti razvoj naziva hrvatskog jezika, od ilirskoga, preko slovinskoga do hrvatskoga.

staroslavenskom jeziku. Godine 1876. odobren mu je izlazak iz franjevačkog reda trećeredaca i imenovan je kanonikom u Zavodu sv. Jeronima u Rimu. Umro je u Rimu na Božić 1902. godine.

3.9.2. Rječnik hrvatsko-talijanski. Vocabolario croato-italiano

Rječnik: *Rječnik hrvatsko-talijanski* tiskan je u Zadru 1901. kao treće popravljeno i povećano izdanje hrvatsko-talijanskog rječnika nakladom *Narodnog lista*. Prvo izdanje pod naslovom *Rječnik ilirsko-talijanski* tiskano je 1858. godine u Zadru u nakladi Petra Abelića, dok je drugo izdanje pod naslovom *Rječnik slovinsko-talijanski* tiskano u Zadru 1874. godine u nakladi braće Battara. Pretisak trećeg izdanja dvojezičnika izašao je 1995. godine. Već iz naslova izdanja rječnika moguće je pratiti razvoj naziva hrvatskoga jezika, od ilirskoga, preko slovinskoga do hrvatskoga.

Budući da je bio na tragu zagrebačke filološke škole, njegova rječnička građa odražava stanja hrvatskoga leksičkog korpusa prije prevlasti vukovske struje u hrvatskome jezikoslovlju. Leksik za rječnike prikupljao je iz narodnog jezika obilazeći Krk, Zadar, kopno i otoke sam ili uz pomoć drugih.⁴⁸ Osim riječi koje se upotrebljavaju u najširim narodnim slojevima, unosi leksik iz područja znanosti, tehnike, zakonodavstva i književnosti, kako bi se u javni život Dalmacije mogao uvesti narodni jezik. Kad je u pitanju stručna i znanstvena terminologija, Parčićev uzor bio je Bogoslav Šulek. Parčić je smatrao da moderan rječnik treba biti bogat kako bi odgovorio na potrebe većine

⁴⁸U prvom izdanju hrvatsko-talijanskog rječnika *Rječnik ilirsko-talijanski* iz 1858. godine spominje Antuna Berčića, Ivana Berčića i Ivana Belamarića kao prinosnike narodnog govora.

čitatelja koji u njemu traže nepoznate riječi, bilo da se radi o narodnom izrazu ili o stručnom terminu.

Za izradu rječnika Parčić se poslužio starijim leksikografskim djelima: Mikaljinim rječnikom iz 1649., prvim i drugim izdanjem della Bellinog rječnika iz 1728. i 1785. godine, te Stullijevim Rjeccsoslòxjem iz 1806. godine. Osim starijih leksikografskih djela, koristio je i suvremene rječnike, pa tako spominje Drobnićev rječnik izdavan u Beču od 1846. do 1849., Fröhlichov rječnik (vjerojatno misleći na onaj veći iz 1853. – 1854.) te Karadžićev rječnik tiskan 1852. u Beču. Mažuranić-Užarevićev rječnik iz 1842. ne spominje. Naime, taj rječnik nije imao hrvatski polazni stupac, pa time nije bio prikladan za korištenje. Osim toga, Fröhlichov rječnik ionako je sadržavao gotovo čitavu leksičku građu Mažuranić-Užarevićevog rječnika.

Treće izdanje Parčićeva hrvatsko-talijanskog rječnika bogato je novotvorenicama i prevedenicama, ali i riječima prikupljenim u narodu. Rječnik je također bogat posuđenicama, koje su precizno označene⁴⁹. Ukupno ima 74 128 natuknica.

Parčićovo zanimanje za botaniku odrazilo se i na njegove rječnike. Vajs⁵⁰ je usporedila Parčićev rječnik iz 1901. s našim najvećim fitonimijskim korpusom, Šulekovim *Jugoslavenskim imenikom bilja* iz 1879. godine. Otkriveno je da se jedna petina Parčićevih potvrda fitonima ne nalazi u Šuleka, što je dokaz da je Parčić bio veliki poznavatelj bilja. U duhu purističkog pristupa

⁴⁹Turk, Marija. "HRVATSKI PRIJEVODNI EKVIVALENTI STRUČNOG NAZIVLJA U PARČIĆEVIM RJEČNICIMA." *FLUMINENSIA: Časopis Za Filološka Istraživanja* 18, no. 1 (Studen 2006): 41-51.

⁵⁰Vajs, Nada. "FITONIMIJA U PARČIĆEVU RJEČNIKU HRVATSKO-TALIJANSKOME IZ 1901." *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje* 29, no. 1 (December 10, 2003): 339-53.

jeziku, posebnu pažnju posvetio je sakupljanju narodnih naziva i njihovoj identifikaciji. U rječniku možemo naći oko 1 670 fitonima, što je zaista velika zastupljenost s obzirom na činjenicu da se radi o općem dvojezičniku.

Parčićev rječnik iz 1901. sastoji se od tri naslovne stranice (str. III.-V.), dvije stranice predgovora drugom izdanju (str. VII.-VIII., *Predgovor k drugomu Izdanju*), dvije stranice predgovora trećem izdanju (str. IX.-X., *Predgovor k trećemu Izdanju*), dvije stranice popisa kratica i odrednica upotrijebljenih u rječniku (str. XI.-XII., *Tablica pokraćenih rieči. (Spiegazione delle abbreviature)*) te 1 237 stranica rječnika.

U *Predgovoru k drugomu Izdanju* Parčić navodi da je u ovo izdanje unio sve natuknice iz svog talijansko-hrvatskog rječnika iz 1868. godine te da je ispravio pogreške iz prvog izdanja iz 1858. godine. Također je izdanje nadopunio novim riječima i izrazima iz narodnog jezika, koje je prikupio sam ili uz pomoć prijatelja. Spominje dva prijatelja koja su mu pomogla u prikupljanju. Jedan je kanonik Mihovil Pavlinović koji mu je iz vlastite zbirke prikupljenih riječi ustupio „znameniti broj od više tisuća rieči“. Te su riječi označene kraticom *P*. Drugi prijatelj je prečasni kanonik J. Mrkica čije je riječi i primjedbe unio od slova L do Ž. Ne navodi da su ove riječi posebno označene. Parčić se nada da je takvim postupkom poboljšao rječnik, a time olakšao učećoj mladeži opširnije poznavanje hrvatskog jezika. Slijedi kratka rasprava o pravopisu u kojoj spominje kako je „ponajviše slijedio etimologički način pisanja, ali se obazrio kadikad i na blagoglasni ili eufonički, gdje god je ovaj pretežnije mah preuzeo“. Služio se dakle umjerenom etimologijom. Odlučio je da neće svaku riječ pisati na oba načina, jer bi time djelo postalo preopsežno te nedovoljno korisno. Osim toga, navodi da koristi slogotvorno *r*, tako da ne piše *kèrv*, već *krv*, uz obrazloženje da glavna suvremena djela pripisuju slovu *r* i

samoglasno svojstvo. U nastavku raspravlja o problemu pronalaska riječi koje se izvorno pišu s *ie*, a njihove izvedenice s *je* ili s *e*. Jedan od primjera koji navodi je *grieh*, *grešnik*, *grješnik*. Savjetuje čitatelja da „se potruži zaviriti i tamo i amo, da nijesmo gdjegod ili ja ili on krivo shvatili one male razlike, kojih još ni svi današnji pisci potanko ne razabiru, il jih nijesu učenjaci još toč[!]no označ[!]ili“. Na kraju piše zahvale te bilježi da je predgovor napisan u Zadru na Cvjetnicu 1874. godine.

U *Predgovoru k trećemu Izdanju* Parčić spominje da je u ovom izdanju ispravio pogreške prošlog, posebno u talijanskom dijelu. Osim toga, nadopunio ga je s više od 15 000 novih riječi iz područja pučkog govora te školskih i poučnih knjiga. Zatim slijedi rasprava o pravopisu, u kojoj se Parčić protivi novoj struji hrvatskih vukovaca koja se zalaže za fonološki pravopis. S njihovim načelom „piši kako govoriš“ Parčić se ne slaže, već smatra da je njegovo načelo pravilno: „piši za oko, a govori za uho“. On se drži umjerenog etimološkog pravopisa, iako nije bio u upotrebi u školama. Ne prihvaća nova (Daničićeva) slova: *đ*, *l*, *ń*, *ǵ*. Također je nezadovoljan uvođenjem novih sinkretskih množinskih padeža⁵¹ umjesto „lijepoga bogatstva svojih padeža“ te se boji da ne dođe do osiromašenja jezika, kao što se dogodilo s bugarskim jezikom. Nakon rasprave o pravopisu, Parčić se nada da će njegovo četrdesetogodišnje iskustvo u sastavljanju rječnika doći do izražaja u ovom izdanju i da će biti korisno učećoj mladeži te odraslima koje zanima poznavanje hrvatskog i talijanskog jezika. Na kraju bilježi da je predgovor napisan u Rimu na Ćirilovo 1901. godine.

U zadnjem dijelu, prije same rječničke građe, nalazi se popis kratica i odrednica upotrijebljenih u rječniku (*Tablica pokraćenih rieči. (Spiegazione delle abbreviature)*). Objašnjenje kratica je na talijanskom jeziku, bez

51 Napušteni su stari padežni oblici za dativ, lokativ i instrumental množine.

hrvatskoga prijevoda. U Tablici 3.1 nalaze se izvorne kratice, njihovo izvorno objašnjenje na talijanskom te prijevod na hrvatski jezik.

Tablica 3.1 Tablica kratice s prijevodom

izvorna kratica	izvorno objašnjenje	prijevod
agg.	aggettivo.	pridjev
(anat.)	termine di anatomia.	anatomija
(antic.)	voce della lingua slava antica.	staroslavenski
(arch.)	t. di architettura.	arhitektura
(arit.)	t. di aritmetica.	aritmetika
(astr.)	t. di astronomia.	astronomija
aum.	aumentativo.	uvećanica
avv.	avverbio.	prilog
(bot.)	t. di botanica.	botanika
(cech.)	t. della lingua čeha.	češki
(chim.)	t. di chimica.	kemija
(chir.)	t. di chirurgia.	kirurgija
(coll.)	nome collettivo.	zbirna imenica
(comm.)	t. commerciale.	trgovina, trgovački
cong.	coniunzione.	veznik
dat.	dativo.	dativ
deriv.	derivati.	izvedenica
dim.	diminutivo,	umanjenica
(dipl.)	t. di diplomazia.	diplomacija
(eccl.)	t. ecclesiastico.	crkveno
f.	femminile.	ženski rod
(fam.)	nel discorso familiare.	neformalni govor
(farm.)	t. di farmacia.	farmacija
(fig.)	nel parlar traslato o metaforico.	preneseno značenje ili metafora
(fil.)	t. filologico.	jezikoslovlje
(filos.)	t. di filosofia.	filozofija
(fis.)	t. di fisica.	fizika
fpl.	femminile plurale.	ženski rod množine
gen.	genitivo.	genitiv
(geogr.)	t. di geografia.	zemljopis

izvorna kratica	izvorno objašnjenje	prijevod
(geom.)	t. di geometria.	geometrija
(gr.)	t. di lingua greca.	grčki
(gram.)	t. di grammatica.	gramatika
impers.	impersonale.	neosoban
[!])indecl.)	indeclinabile,	nesklonjivost, indeklinabilnost
[!])nterj.	interjezione.	uzvik
(it.)	voce italiana.	talijski
(leg.)	t. legale o di giurisprudenza.	pravni termin
(log.)	t. di logica.	logika
m.	maschile.	muški rod
(mag.)	t. di lingua magiara.	mađarski
(mar.)	t. di marineria.	pomorstvo
(mat.)	t. di matematica.	matematika
[!])mec.)	t. di meccanica.	mehanika
(med.)	t. di medicina.	medicina
(mere.)	t. di mercatura.	tržište, financije, trgovinska razmjena
(mil.)	t. militare.	vojnički izraz
(miner.)	t. di mineralogia.	mineralogija
(mitol.)	t. di mitologia.	mitologija
mo. avv.	modo avverbiale.	priložna oznaka
(mor.)	in senso morale.	moralne vrijednosti
(mus.)	t. di musica,	glazba
mpl.	maschile plurale.	muški rod množine
n.	neutro.	srednji rod
npl.	neutro plurale,	srednji rod množine
num.	numerale.	broj
(P)	t. della raccolta Pavlinović.	Pavlinovićeva zbirka
part, partic.	particella,	čestica
partep.	participio.	glagolski pridjev radni/trpni ili glagolski prilog sadašnji/prošli
(pitt.)	t. di pittura.	slikarstvo, likovna umjetnost
pl.	plurale.	množina
(poet.)	t. di poesia o delle canzoni nazionali.	poezija ili narodne pjesme
(pol.)	t. di politica,	politika
prep.	preposizione,	prijedlog

izvorna kratica	izvorno objašnjenje	prijevod
pron.	pronome.	zamjenica
(prov.)	proverbio,	poslovica
qc.	qualche cosa,	nešto
qd.	qualcheduno.	netko
(ret.)	rettorica.	retorika
(russ.)	voce russa.	ruski
(scherz.)	voce da scherzo,	pošalica, šala
seg.	seguenti.	sljedeći
sprezz.	t. sprezzativo.	pejorativ
(slov.)	t. sloveno.	slovenski
(ted.)	voce tedesca.	njemački
(teol.)	t. di teologia.	teologija
(tit.)	ne' titoli.	titula
V.	Vedi.	vidi
vai.	verbo attivo imperfettivo.	prijelazni nesvršeni glagol
vap.	verbo attivo perfettivo.	prijelazni svršeni glagol
(veter.)	t. di veterinaria,	veterina
vezz.	vezzeggiativo,	hipokoristik
vni.	verbo neutro imperfettivo,	neprijelazni nesvršeni glagol
vnp.	verbo neutro perfettivo.	neprijelazni svršeni glagol
(volg.)	t. triviale o volgare,	trivijalno ili vulgarno
vri.	verbo riflessivo imperfettivo,	povratni nesvršeni glagol
vrp.	verbo rifiess. perfettivo.	povratni svršeni glagol
*	voce turca.	turski

Rječnik sadrži posuđenice iz staroslavenskog, češkog, grčkog, talijanskog, mađarskog, ruskog, slovenskog, njemačkog, turskog, latinskog i francuskog jezika. U *Tablici pokraćenih rieči* nedostaju kratice za latinski i francuski jezik. U Tablici 3.2 može se vidjeti popis posuđenica te broj pojavljivanja označenih posuđenica u rječniku. Usprkos činjenici da u ono doba

nisu postojali etimološki priručnici za hrvatski jezik, Parčić je uspješno identificirao podrijetlo većine posuđenica.⁵²

Tablica 3.2 Tablica posuđenica

izvorna kratica	izvorno objašnjenje	prijevod	broj potvrda
*	voce turca.	turski	1 468
(it.)	voce italiana.	talijanski	303
(ted.)	voce tedesca.	njemački	99
(mag.)	t. di lingua magiara.	mađarski	98
(antic.)	voce della lingua slava antica.	staroslavenski	93
(gr.)	t. di lingua greca.	grčki	64
(slov.)	t. sloveno.	slovenski	28
(russ.)	voce russa.	ruski	26
(lat.)		latinski	4
(franc.) / (frances.)		francuski	4
(cech.) [(čech.)]	t. della lingua čeha.	češki	1

U rječniku nalazimo uglavnom samo riječi, ali možemo naći i sintagmatske izraze i frazeološke veze.

Strukovno i znanstveno nazivlje naznačeno je kraticama. U Tablici 3.3 možemo vidjeti koja su područja u rječniku pokrivena.

52 Turk, Marija. "Stranojezični Elementi U Rječnicima Dragutina Parčića." *Filologija*, no. 59 (Ožujak 2013): 195-205.

Tablica 3.3 Tablica strukovnog i znanstvenog nazivlja

izvorna kratica	izvorno objašnjenje	prijevod	broj potvrda
(anat.)	termine di anatomia.	anatomija	286
(arch.)	t. di architettura.	arhitektura	28
(arit.)	t. di aritmetica.	aritmetika	1
(astr.)	t. di astronomia.	astronomija	58
(bot.)	t. di botanica.	botanika	178
(chim.)	t. di chimica.	kemija	136
(chir.)	t. di chirurgia.	kirurgija	9
(comm.)	t. commerciale.	trgovina, trgovački	79
(dipl.)	t. di diplomazia.	diplomacija	3
(eccl.)	t. ecclesiastico.	crkveno	101
(farm.)	t. di farmacia.	farmacija	1
(fil.)	t. filologico.	jezikoslovlje	8
(filos.)	t. di filosofia.	filozofija	17
(fis.)	t. di fisica.	fizika	192
(geogr.)	t. di geografia.	zemljopis	13
(geom.)	t. di geometria.	geometrija	107
(gram.)	t. di grammatica.	gramatika	156
(leg.)	t. legale o di giurisprudenza.	pravni termin	345
(log.)	t. di logica.	logika	2
(mar.)	t. di marineria.	pomorstvo	60
(mat.)	t. di matematica.	matematika	84
[!])mec.)	t. di meccanica.	mehanika	21
(med.)	t. di medicina.	medicina	229
(mere.)	t. di mercatura.	tržište, financije, trgovinska razmjena, knjigovodstvo?	0
(mil.)	t. militare.	vojnički izraz	86
(miner.)	t. di mineralogia.	mineralogija	12
(mitol.)	t. di mitologia.	mitologija	15
(mus.)	t. di musica,	glazba	57
(pitt.)	t. di pittura.	slikarstvo, likovna umjetnost	15
(poet.)	t. di poesia o delle canzoni nazionali.	poezija ili narodne pjesme	413
(pol.)	t. di politica,	politika	9
(ret.)	rettorica.	retorika	11
(teol.)	t. di teologia.	teologija	79

izvorna kratica	izvorno objašnjenje	prijevod	broj potvrda
<i>(veter.)</i>	t. di veterinaria,	veterina	2

4. Digitalizacija rječnika

4.1. Opća načela digitalizacije

Danas u svijetu postoji nebrojeno mnogo rječnika i enciklopedija u elektroničkom obliku, bilo u tekstualnom obliku bilo u strukturiranom obliku rječničke (enciklopedijske) baze podataka s različitim indeksima, mogućnostima pretraživanja i slično.

Međutim, do prije deset godina, dakle do 2004. godine, gotovo da nije ni bilo starijih rječničkih tekstova u elektroničkom obliku.

Suvremeni su elektronički rječnici sustavi zatvoreni u sebe i ne omogućuju međusobno povezivanje niti kritičku usporedbu jer su pretežno trgovačke prirode, ali jednako i u slučajevima kad omogućuju slobodan pristup, taj pristup ograničen je sustavom kroz koji se pristupa podacima i koji ne omogućuje preuzimanje podataka.

Istovremeno, postoji veoma mnogo projekata kojima je cilj objavljivanje stare tekstualne građe (poput projekta *Gutenberg* ili *Titus*), objavljivanja biblijskih tekstova, književnih djela, pa i rječnika, ali do sada za hrvatski jezik rječnici na taj način nisu obrađivani. Tako na primjer sustav Gutenberg, koji je prvi takav sustav u svijetu, nastao 1971. godine, te omogućuje pristup velikom broju digitaliziranih knjiga (e-knjiga) u tekstualnom formatu, iako kad god je moguće omogućuje i pristup knjigama u HTML, PDF te nekim drugim formatima za mobilne telefonske uređaje, u sebine sadrži ni knjige ni rječnike,

recimo, na hrvatskom jeziku.⁵³ Projekt Gutenberg nije ni bio zamišljen kao sredstvo koje bi pomoglo znanstvenicima u obradi i analizi starih povijesnih dokumenata ili tekstova rječnika, nego je usmjeren isključivo na opću publiku te sadrži većinom beletrističke tekstove. Stoga je ovdje naveden isključivo zbog povijesnih razloga budući da je bio prvi takav sustav.

S druge strane, projekt Titus namijenjen je suradnji među znanstvenicima koji se bave indoeuropskim jezicima i omogućuje pristup svim materijalima projekta. Projekt *Titus* (na njemačkom Thesaurus Indogermanischer Text- und Sprachmaterialien – Tezaurusindogermanskih tekstova i govora) važan je projekt Goetheovog sveučilišta u Frankfurtu na Majni⁵⁴ kojemu je cilj da prikuplja informacije o indoeuropskim jezicima te da poboljšava suradnju među stručnjacima i znanstvenicima. Projektu⁵⁵ je „cilj da bude pomoć u istraživanjima vezanim na računalne obrade tekstualnih i jezičnih podataka te da prikuplja rječnike, popise riječi, kao i alate za lingvističke analize.

Svima onima koji doprinose projektu s određenim tekstualnim prilogom ili materijalom, odnosno programom omogućuje se pristup podacima, a nekim

⁵³ Iako projekt Gutenberg sadrži preko 45 tisućak knjiga na velikom broju jezika, nema uopće knjiga na hrvatskom jeziku. Postoji samo nekoliko tekstova na slovenskom (1 tekst) i srpskom (4 teksta) (kao na manje poznatim jezicima), međutim, niza je jezika koje je uključio u povijesne tekstove. Sustav je ograničen pretežitom na engleski jezik, te je trenutno pristupa imao 38 687 engleskih digitalnih tekstova, pri čemu je još samo 15 jezika imalo više od 50 tekstova, a samo je od 50 tekstova imao još 49 jezika, a najčešće postoje samo jedan tekst što se ukupno obuhvaća prema objavljenom podatku točno 46 654 e-knjiga slobodnim pristupom. <http://www.gutenberg.org/catalog/> (31. 8. 2014.).

⁵⁴ Johan Wolfgang Goethe Universität, Frankfurt am Main. Thesaurus Indogermanischer Text- und Sprachmaterialien <http://titus.uni-frankfurt.de/indexe.htm>. (31. 8. 2014.).

⁵⁵ Prema članku Thesaurus Indogermanischer Text- und Sprachmaterialien u Wikipediji http://de.wikipedia.org/wiki/Thesaurus_Indogermanischer_Text-_und_Sprachmaterialien te je zajednički projekt odsjeka za lingvistiku s više sveučilišta pokrenut 1987.

datotekama može se pristupati slobodno. Načini prikaza rječničkih podataka i ostalih tekstova analizirat će se u sljedećim poglavljima.“

Današnji elektronički rječnici većinom su tek preoblika klasičnih papirnatih rječnika kojima je dodano pretraživanje punoga teksta i međusobno povezivanje natuknica na temelju pojave ključnih riječi u tekstu natuknice, ali nisu dodane strukture koje bi rječnike pretvorile u izvore znanja. Na enciklopedijskoj razini, u velikim svjetskim elektroničkim enciklopedijama poput Britannice (pa i Encarte) dodano je strukturiranje tekstova prema temama i selektiranje natuknica prema općim ključnim riječima (kategorijama), ali za jednojezične rječnike koji su također dio tih enciklopedija strukturiranje je ostvareno tek na razini vrste riječi.

Upravo su to bili razlozi zašto je pokrenut projekt „Hrvatska rječnička baština i prikaz rječničkoga znanja“⁵⁶(kao što je istaknuto u prikazu toga projekta). Budući da „današnje enciklopedije na novim medijima mogu u tom smislu predstavljati izvor znanja jer mu pristupaju kroz određene zadane strukture, rječnici to još nisu jer još nema znanstvenoga aparata koji bi kategorizaciju riječi prema upotrebi, značenju, nastanku itd. primijenio u konkretnoj izvedbi.“

U strojnoj obradi prirodnoga jezika temeljenoj pretežno na korpusima, označivanje riječi kao jedna od najvažnijih zadaća temelji se većinom na različitim morfološkim algoritmima, a rječničke i jezične baze podataka gotovo

⁵⁶Boras, Damir. **Croatian dictionary heritage - collection of old Croatian dictionaries in digital form.**Conference on Journalism and Information Technology and UNESCO project on the word heritage Colloquium on the project "Croatian dictionary heritage and dictionary knowledge presentation" Inter-University Centre, Dubrovnik, May 26-28, 2004 (pozvano predavanje, neobjavljeni rad) <http://bib.irb.hr/prikazi-rad?&rad=182779>.

da se i ne upotrebljavaju iz jednostavnoga razloga što ih nema odnosno nisu dostupne.

Što se tiče starijih hrvatskih dvojezičnih i višejezičnih rječnika, oni su do sada bili povremeno i djelomično obrađivani ručnim metodama, od kojih nijedna nije mogla biti sigurna u sveobuhvatnost obrade. Čak ni objavljivanje pretisaka nekoliko najvažnijih starih hrvatskih rječnika u posljednjih 25 godina nije značajno povećalo znanstvenu literaturu o njima.

Tek se uvidom i pažljivom analizom može primijetiti da su neki rječnici poput Vrančićeva, Mikaljina i Belostenčeva u većoj ili manjoj mjeri utjecali i poslužili kao građa za današnje još uvijek jedine moderne jednojezične hrvatske rječnike (Novi Liber te Leksikografski zavod-Školska knjiga) ili za Osmojezični enciklopedijski rječnik Leksikografskoga zavoda. Stoga sve više raste potreba za izradom korpusa tekstova hrvatske rječničke baštine na načelima strukturiranja rječničkog znanja uz pomoć informacijske tehnologije.⁵⁷

U bibliografskom smislu postoji dosta dobra obrađenost hrvatske rječničke baštine, ali kao tematska sveobuhvatna cjelina nije do sada sastavljena jer se bibliografska obrada rječničke građe obuhvaćala uglavnom kroz bibliografije rijetkih starih hrvatskih knjiga, ili kao dio postojeće građe pojedinih biblioteka, tako da je do sada rječnička baština bila obuhvaćana uglavnom djelomično.

Također, u novije se vrijeme po hrvatskim arhivima i samostanima otkrivaju različiti stari rukopisni rječnici i građa koja do sada nije obrađena, pa je potreba za takvom bibliografijom i za istraživanjem nepoznate hrvatske rječničke građe također velika.

⁵⁷Boras, Damir. Op. cit.

Iako su se mnogi koristili tom građom, nijedan od tih starijih rječnika do sada nije strojno obrađen i nije u potpunosti analizirana njihova struktura.

U tehničkom smislu, kroz projekte *Hrvatska rječnička baština i prikaz rječničkoga znanja* i *Hrvatska rječnička baština i hrvatski europski identitet* u dobrom su dijelu ostvareni ciljeva projekta koji se odnose na izradu strojne baze tekstova najvažnijih starijih hrvatskih dvojezičnih i višejezičnih rječnika kojom se u načelu omogućuju sveobuhvatna rječnička, jezična i gramatička analiza, zatim razne terminološke analize rječničkih tekstova, kao i usporedne i temporalnu analize između pojedinih rječnika te, što je najvažnije, omogućuju i istraživanja strukture rječničkoga znanja sadržanoga u tim rječnicima.

Time su stvorene pretpostavke da se, na temelju stečenih iskustava, znanstvenom analizom utvrde načela strojne pripreme rječnika koje se tiču digitalizacije slike, upisa teksta, transkripcije te prijevoda na današnji hrvatski jezik itd.

Također, bilo je potrebno utvrditi načela prikaza znanja sadržanoga u tim rječnicima kako bi se omogućila daljnja usporedna istraživanja ne samo sa suvremenim nego i s današnjim rječnicima te potvrdila temeljna hipoteza projekata da se upotrebom informacijske tehnologije rječnici mogu tako strukturirati da postanu izvor novoga znanja, i ne budu samo popisi riječi nego i izvori, npr. terminološkoga, etimološkoga i drugoga znanja.

Kada je riječ o načelima strojne pripreme rječnika koje se tiču digitalizacije slike, upisa teksta, transkripcije te prijevoda na današnji hrvatski jezik, prihvaćena su načela da pri svim tim aktivnostima treba izgubiti što je manje informacija sadržanih u izvornim predlošcima rječnika, ili općenito govoreći u obrađivanim dokumentima.

Za **digitalizaciju slike** utvrđeno⁵⁸ je da to mora biti digitalizacija u boji (24 – 48bita), da rezolucija mora biti oko 400 dpi, te da pohrana digitalne slike mora biti u obliku kojim se ne gubi informacija o slici (npr. TIFF format). Za prikaz u javnoj domeni putem mreže (interneta) prihvatljivi su i formati kojima se djelomično gubi snimljena informacija, ali je kvaliteta i dalje prihvatljiva (tome npr. odgovora JPG format).

Za **zapis teksta** prihvaćeno⁵⁹ je načelo doslovnog prijepisa teksta što znači da se prepisuju i sve pogreške u izvornom slogu teksta rječnika (dokumenta). Pri tome treba upotrebljavati i font koji najbliže odgovara izvornom tisku veličinom, odnosom visine i širine slova te bojom. Potrebno je zadržati vezu prema izvorno tiskanim stranicama, te redcima sloga. Poželjno je da digitalni tekst bude oblikovan tako da u potpunosti oponaša izgled, veličinu i raspored teksta na izvorno tiskanom slogu.

U pogledu **transkripcije**⁶⁰ na današnji pravopis radi omogućivanja pretraživanja teksta prihvaćeno je načelo jednoznačne transliteracije teksta kako bi se omogućilo doslovno automatsko preobličavanje na modernu grafiju, ali isto tako i omogućila naknadna ručna prilagodba na suvremeni pravopis, kao što su to načinili Boras i Mikelić na primjeru Vrančićeva rječnika.⁶¹

Jednako treba ostaviti mogućnost da se u digitalizirani tekst doda i **prijevod na današnji hrvatski jezik** kako bi se omogućile dodatne analize i

⁵⁸Boras, Damir. Op. cit.

⁵⁹Boras, Damir. Op. cit.

⁶⁰Boras, Damir. Op. cit.

⁶¹**Boras, Damir; Mikelić, Nives.**Rječnik Fausta Vrančića – temelj hrvatske rječničke baštine (računalna obradba)// Modeli znanja i obrada prirodnoga jezika / Tuđman, Miroslav (ur.). Zagreb : Zavod za informacijske studije, Filozofski fakultet, 2003. Str. 237-272.

obrade⁶² te kako bi se na taj način, uz utvrđena opća načela strojne pripreme starije (pa i suvremene) leksikografske građe, znanstvenoj publici (ne samo domaćoj) s područja leksikografije, enciklopedike, leksikologije, lingvistike, kroatologije i ostalih znanosti (na primjer medicine – terminološka istraživanja) pružio korpus za razna buduća istraživanja.

Stoga ćemo u sljedeća dva poglavlja na nekoliko najvažnijih primjera analizirati kakva su svjetska, europska te hrvatska iskustva u digitalizaciji povijesnih tekstova, prvenstveno rječničkih, te kako i u kojem obliku omogućuju pristup rječničkim odnosno povijesnim tekstovima.

4.2. Svjetska i europska iskustva

U svijetu, kao što je djelomično pregledno prikazano u prethodnom poglavlju, postoji velik broj internetskih stranica koje omogućuju pristup izvornim tekstovima različitih dijela, uključujući tu i rječnike, među kojima ima i povijesnih.

S razvojem tehnologije povećao se broj projekata koji na sličan način omogućuju preuzimanje ili upotrebu digitaliziranih tekstova, pri čemu će se značajne projekte poput projekta Gutenberg, Titus,⁶³ Perseus⁶⁴ i neke druge posebno analizirati u ovom poglavlju. Budući da je gotovo nemoguće nabrojati sve moguće stranice te vrste, pa tako na primjer samo u svojoj knjizi Piotrowski⁶⁵ u pregledu navodi po svojem izboru 54 korpusa povijesnih tekstova

⁶²Boras, Damir; Mikelić, Nives. Op. cit.

⁶³Projekti Gutenberg i Titus, vidi bilješke 31, 32 i 33.

⁶⁴Perseus Digital Library, Tufts University. <http://www.perseus.tufts.edu/hopper/> (2. 9. 2014.).

⁶⁵Piotrowski, op. cit, str. 101.

na više od deset različitih jezika (Arapski – 2, Kineski – 3, Nizozemski – 5, Engleski – 13, Francuski – 7, Portugalski – 5, Njemački – 8, Nordijski jezici – 4, Latinski i stari Grčki – 7), u ovom poglavlju ukratko ćemo prikazati i analizirati nekoliko odabranih.

Svrha je tih prikaza i analiza da se pokaže u kojem smjeru treba razviti model prezentacije starijih hrvatskih rječnika.

4.2.1. Projekt *Gutenberg*

Projekt Gutenberg praktički je prvi projekt digitalizacije, pokrenut 1971. godine, te je početno bio instaliran na *mainframe* računalu Sveučilišta Illinois, koje je bilo jedan od samo 15 čvorova računalne mreže ARPANET koja je kasnije postala internet.⁶⁶ Ovdje se projekt Gutenberg spominje prvenstveno zbog navedenih povijesnih činjenica, jer, kao što je već spomenuto, nije ni bio zamišljen kao pomagalo znanstvenicima, nego je usmjeren isključivo na opću publiku te sadrži većinom beletrističke tekstove.

Prema najnovijim podacima, objavljenima na internetskim stranicama projekta Gutenberg (iz svibnja 2014.), tvrdi se da zbirka sadrži preko 45 tisuća digitaliziranih jedinica, te da raste po prosječnoj stopi od 50 novih e-knjiga tjedno. Riječ je prvenstveno o literarnim djelima zapadne kulturne tradicije, pretežno novelama, poeziji, kratkim pričama i dramama u tekstualnom obliku, iako ima i ponešto zvučnih i glazbenih notnih zapisa.⁶⁷

⁶⁶Sastavljenopremapodacimaiz

člankauWikipedijipodnaslovom*Project Gutenberg*http://en.wikipedia.org/wiki/Project_Gutenberg (25.8. 2014.).

⁶⁷Ibid.

Digitalizirani tekstovi doprinose se isključivo na volonterskoj osnovi, a građa se objavljuje u najjednostavnijem tekstualnom formatu, kodirana na najjednostavniji način. Riječi je o najčešće o najjednostavnijem US-ASCII kôdu, ali se objavljuju i tekstovi kodirani prema ISO-8859-1 standardu koji omogućuje upis znakova za jezike zapadne Europe. U novije se vrijeme tekstovi, naročito oni na kineskom ili japanskom, kodiraju prema UTF-8 standardu. Ako se kroz volonterski doprinos pojave, objavljuju se i inačice tekstova i u složenijim formatima poput HTML formata koji omogućuju i objavljivanje slika. Složeniji formati poput PDF formata koji ne omogućuju jednostavno ispravljanje podataka ne smatraju se sukladnim ciljevima projekta.⁶⁸

Zanimljivo je napomenuti u kontekstu ovoga rada da projekt ima dva glavna formata u kojima se pohranjuju svi podaci, koji su prihvatljivi kao formati volonterskih doprinosa i iz kojih se generiraju sve ostale vrste datoteka. Jedan format je prilagođena verzija TEI (*Text Encoding Initiative*)⁶⁹ standarda koja se počela upotrebljavati 2005. godine, a drugi oblik je označen prema standardu reStructuredText⁷⁰ koji se počeo upotrebljavati 2011. godine.⁷¹

Važno je napomenuti da se sve e-knjige dostupne preko projekta Gutenberg mogu slobodno koristiti te na njima nema autorskih prava (*copyright*).

Rječničke digitalizirane građe gotovo da i nema, te se, na primjer, u objavljenom digitalnom katalogu⁷² (koji je također objavljen u jednostavnom

⁶⁸Ibid.

⁶⁹“TEI: Text Encoding Initiative.” <http://www.tei-c.org/index.xml> (19. 2. 2013.).

⁷⁰Wikipedija, članak *reStructuredText* <http://en.wikipedia.org/wiki/ReStructuredText> (30. 8. 2014.).

⁷¹Premanatuknici *Project Gutenberg* http://en.wikipedia.org/wiki/Project_Gutenberg (25. 8. 2014.).

⁷²*Off-line* katalog Projekta Gutenberg <http://www.gutenberg.org/dirs/GUTINDEX.ALL> (2. 9. 20014.).

tekstualnom obliku) od preko 36 tisuća naslova na engleskom jeziku može naći samo 99 referenci koje sadrže riječ *dictionary*, 17 referenci sadrži riječ *vocabulary*, 3 reference sadrže riječ *thesaurus*, 1 referenca sadrži riječ *Wörterbuch*, 4 reference sadrže riječ *dizionario* (i te se odnose na 4 sveska jednog te istoga rječnika), a 1 referenca sadrži riječ *vocabulaire*. Kao što je već spomenuto, građe na hrvatskom jeziku, pa tako ni rječničke, uopće nema.

Dostupna rječnička građa dostupna je najčešće u HTML obliku, a katkada u tekstualnom obliku koji sadrži HTML oznake. To se najbolje vidi na primjeru teksta engleskoga rječnika *Chambers's Twentieth Century Dictionary* iz 1908. godine.⁷³

Primjer uvodnog teksta rječnika:

The Project Gutenberg EBook of Chambers's Twentieth Century Dictionary
(part 1 of 4: A-D), by Various

This eBook is for the use of anyone anywhere at no cost and with almost no restrictions whatsoever. You may copy it, give it away or re-use it under the terms of the Project Gutenberg License included with this eBook or online at www.gutenberg.net

Title: Chambers's Twentieth Century Dictionary (part 1 of 4: A-D)

Author: Various

Editor: Thomas Davidson

Release Date: October 9, 2011 [EBook #37683]

Language: English

Character set encoding: ISO-8859-1

*** START OF THIS PROJECT GUTENBERG EBOOK CHAMBERS'S 20TH CENT. DICTIONARY. (A-D)

Produced by Keith Edkins, Colin Bell and the Online
Distributed Proofreading Team at <http://www.pgdp.net>

⁷³*Chambers's Twentieth Century Dictionary of the English Language*(part 1 of 4: A-D), London and Edinburgh, 1908. <http://www.gutenberg.org/files/37683/37683-h/37683-h.htm> (30.8. 2014.).

Transcriber's note:

A few typographical errors have been corrected. They appear in the text like this, and the explanation will appear when the mouse pointer is moved over the marked passage.

CHAMBERS'S
TWENTIETH CENTURY DICTIONARY
OF THE
ENGLISH LANGUAGE
PRONOUNCING, EXPLANATORY, ETYMOLOGICAL, WITH COMPOUND PHRASES,
TECHNICAL TERMS IN USE IN THE ARTS AND SCIENCES,
COLLOQUIALISMS, FULL APPENDICES, AND
COPIOUSLY ILLUSTRATED
EDITED BY
REV. THOMAS DAVIDSON
ASSISTANT-EDITOR OF 'CHAMBERS'S ENCYCLOPÆDIA'
EDITOR OF 'CHAMBERS'S ENGLISH DICTIONARY'
LONDON: 47 Paternoster Row
W. & R. CHAMBERS, LIMITED
EDINBURGH: 339 High Street
1908

CHAMBERS'S
NEW LARGE TYPE
ENGLISH DICTIONARY
EDITED BY
REV. THOMAS DAVIDSON
Pronouncing, Explanatory, Etymological
1264 pp. Imp. 8vo, cloth, 12/6; hf.-mor., 18/-

"The best one volume dictionary in existence."
W. & R. CHAMBERS, LIMITED, LONDON AND EDINBURGH.

P R E F A C E .

This is the third English Dictionary which the present Editor has prepared, and he may therefore lay claim to an unusually prolonged apprenticeship to his trade. It is surely unnecessary for him to say that he believes this to be the best book of the three, and he can afford to rest content if the Courteous Reader receive it with the indulgence extended to his Library Dictionary, published in the spring of 1898. It is based upon that work, but will be found to possess many serviceable qualities of its own. It is not much less in content, and its greater relative portability is due to smaller type, to thinner paper, and still more to a rigorous compression and condensation in the definitions, by means of which room has been found for many additional words.

Primjer natuknica u rječniku:

Abigail, ab'i-gāl, *n.* a lady's-maid. [From *Abigail*, 1 Sam. xxv.]

Ability, a-bil'i-ti, *n.* quality of being able: power: strength: skill.—*n.pl.* **Abilities**, the powers of the mind. [O. Fr. *ableté* (Fr. *habileté*)—L. *habilitas*—*habilis*, easily handled, from *habēre*, to have, hold. See **Able**.]

Abintestate, ab-in-tes'tāt, *adj.* inheriting the estate of one who died without having made a will. [L. *ab*, from, and **Intestate**.]

Abiogenesis, ab-i-o-jen'es-is, *n.* the origination of living by not-living matter, spontaneous generation.—*adj.* **Abiogenet'ic**—*n.* **Abio'genist**, one who believes in such. [Coined by Huxley in 1870; Gr. *a*, neg., *bios*, life, *genesis*, birth.]

Abject, ab-jekt', *v.t. (obs.)* to throw or cast down or away. [L. *abjicere*, -*jectum*—*ab*, away, *jacere*, to throw.]

Abject, ab'jekt, *adj.* cast away: mean: worthless: cowering: base.—*n.* an outcast.—*ns.* **Abjec'tion**, **Ab'jectness**, a mean or low state: baseness: degradation.—*adv.* **Ab'jectly**. [L. *abjectus*, cast away—*ab*, away, *jacere*, to throw.]

Abjudge, ab-juj', *v.t. (rare)* to take away by judicial sentence. [L. *ab*, from, and **Judge**.]

Abjudicate, ab-jōō'di-kāt, *v.t.* to give by judgment from one to another. [L. *ab*, from, and **Judicate**.]

Abjure, ab-jōōr', *v.t.* to renounce on oath or solemnly: to recant: to repudiate.—*n.* **Abjurā'tion**, official renunciation on oath of any principle or pretension.—*adj.* **Abjur'atory**.—*n.* **Abjur'er**. [L. *ab*, from, *jurāre*, -*ātum*, to swear.]

Abkari, ab-kā'ri, *n.* the manufacture or sale of spirituous liquors: the excise duty levied on such.—Also **Abka'ry**. [Pers.]

Ablactation, ab-lak-tā'shun, *n.* a weaning. [L. *ab*, from, *lactāre*, to suckle—*lac*, *lactis*, milk.]

Ablation, ab-lā'shun, *n.* the act of carrying away: (*geol.*) the wearing away of rock by the action of water.—*adj.* **Ablati'tious**. [L. *ab*, from, *latum*, supine of *ferre*, to bear.]

Ablative, ab'lat-iv, *adj.* used as a noun. The name applied to one of the cases in the declension of nouns and pronouns in the Indo-European languages, retained as in Latin and Sanskrit, or merged in another case, as in the genitive in Greek. Its meaning was to express *direction from* or *time when*.—*adj.* **Ablati'val**. [L. *ablativus*—*ab*, from, *ferre*, *latum*, to take; as if it indicated taking away, or privation.]

Ablaut, ab'lowt, *n. (philol.)* vowel permutation, a substitution of one root vowel for another in derivation, as in *sing*, *sang*, *song*, *sung*, distinct from the phonetic influence of a succeeding vowel, as in the Umlaut. It is especially the change of a vowel to indicate tense-change in strong verbs. [Ger., from *ab*, off, and *laut*, sound.]

Ablaze, a-blāz', *adj.* in a blaze, on fire: gleaming brightly. [Prep. *a*, and **Blaze**.]

Able, ā'bl, *adj.* (comp. **A'bler**; superl. **A'blest**) having sufficient strength, power, or means to do a thing: skilful.—*adj.* **A'ble-bodied**, of a strong body: free from disability, of a sailor, labourer, &c.: robust.—*adv.* **A'bly**. [See **Ability**.]

Ablegate, ab'le-gāt, *n.* a papal envoy who carries the insignia of office to a newly-appointed cardinal.

Ablloom, a-blōōm', *adv.* in a blooming state. [Prep. *a*, on, and **Bloom**.]

Abluent, ab'lōō-ent, *adj.* washing or cleaning by a liquid.—*n.* a medicine which carries off impurities from the system. [L. *abluens*, -*entis*, pr.p. of *abluere*, to wash away—*ab*, from, away, and *luere* = *lavāre*, to wash. See **Lave**.]

Ablution, ab-lōō'shun, *n.* act of washing, esp. the body, preparatory to religious rites: any ceremonial washing, symbolic of moral purification: the wine and water used to rinse the chalice, drunk by the officiating priest.—*adj.* **Ablu'tionary**. [L. *ablutio*—*ab*, away, *luere* = *lavāre*, to wash.]

Abnegate, ab'ne-gāt, *v.t.* to deny.—*ns.* **Abnegā'tion**, renunciation; **Ab'negator**, one who abnegates or renounces. [L. *ab*, away, and *negāre*, to deny.]

Abnormal, ab-nor'mal, *adj.* not normal or according to rule: irregular—also **Abnor'mous**.—*ns.* **Abnormal'ity**, **Abnor'mity**.—*adv.* **Abnor'mally**. [L. *ab*, away from, and **Normal**.]

Kao zaključak, uočljivo je da uvodni dio rječnika sadrži kompletnu referencu, s označenim načinom kodiranja te napomenom transkriptora.

Slog rječnika (uvod, tablice, natuknice) korektno je grafički strukturiran izborom fonta (Times New Roman, *italic*, **bold**) te izborom znakova, tako da je vrlo pogodan za automatsko strukturiranje teksta.

U tekstu postoje poveznice na stranice rječnika, dok su rječničke natuknice prikazane u bloku te nemaju strukturu po izvornim recima.

Slike izvornog rječničkog sloga nisu dostupne.

4.2.2. Projekt TITUS

Projekt TITUS⁷⁴ (na njemačkom „Thesaurus indogermanischer Text- und Sprachmaterialien“ – Tezaurus indogermanske tekstualne i govorne građe) projekt je Sveučilišta Johann Wolfgang Goethe u Frankfurtu na Majni koji vodi prof. dr. Jost Gippert. Projekt je pokrenut 1987. pod imenom „Thesaurus altindogermanischer Texte auf Datenträgern“⁷⁵ (njem. Tezaurus staroindogermanskih tekstova na nosačima podataka), a cilj mu je prvobitno bio da prikuplja informacije o indoeuropskim jezicima jezika, te poboljša i poveća suradnju među stručnjacima i znanstvenicima te na taj načini olakša rad na komparativnoj analizi jezika u lingvističkoj usporedbi i analizi jezika te rekonstrukciji relevantnih tekstova na staroindogermanskim jezicima u digitalnom obliku.

Ciljevi projekta su i da unaprijedi znanstvena istraživanja vezana za računalnu obradu podataka te da dodatno u tu svrhu u digitalnom obliku

⁷⁴Wikipedija, članak TITUS (project). [http://en.wikipedia.org/wiki/TITUS_\(project\)](http://en.wikipedia.org/wiki/TITUS_(project)) (2. 9. 2014.).

⁷⁵**Bunz, Carl-Martin.** Der Thesaurus indogermanischer Text- und Sprachmaterialien (TITUS) – ein Pionierprojekt der EDV in der Historisch-Vergleichenden Sprachwissenschaft. U: Sprache und Datenverarbeitung 1, 1998, 10-30.

prikuplja osim korpusa tekstova i rječnike, popise riječi te alate za jezične analize. Suradnici su svi oni koji doprinesu projektu s nekom digitalnom građom, te samo oni imaju pristup svim materijalima projekta. Jedan dio digitalne građe javno je dostupan bez naknade.

Projekt omogućuje korisnicima koji se bave nekomercijalnim djelatnostima upotrebu svojega Unicode 4.0 fonta (Titus Cyberbit Basic)⁷⁶, kao i upotrebu odgovarajuće tipkovnice koja sadrži znakove za velik broj jezika koje u indoeuropskim lingvističkim istraživanjima koriste filolozi i lingvisti.

Budući da je TITUS zamišljen kao hipertekstualna baza podataka odnosno zbirka materijala, detaljno su razrađene procedure pripreme digitalnog teksta, koje ćemo ukratko prikazati.⁷⁷

Prvo se izrađuje slika izvornog nosača teksta (ako postoji). Kod zapisa u kamenu, drvu i glini nosač teksta mora se najprije fotografirati. Pri tome treba osigurati dovoljnu rezoluciju slike. Slijedi digitalizacija slike, ako nije odmah korišten digitalni fotoaparat dovoljne rezolucije. Pri tome se također prethodno fotografiraju rukopisi na raznim materijalima poput papira, papirusa ili pergamenta. Kod oštećenih listova svakako treba pripaziti da se pod nosače stavi odgovarajuća podloga kako bi se razlikovala pri digitalizaciji oštećenih površina.

Ako je potrebno, slijedi digitalizacija slike. U današnje vrijeme sve se rješava robotima za automatsko listanje stranica i digitalno automatsko snimanje u dovoljnoj rezoluciji. Zatim slijedi obrada digitaliziranih slika, odnosno obrada

⁷⁶TITUS Unicode <http://titus.uni-frankfurt.de/indexe.htm?unicode/unitest2.htm> (2. 9. 2014.).

⁷⁷PremaBunz, Carl Martin. Ibidem, str. 18.

materijala u jezičnom i filološkom smislu. Pritom treba naglasiti da se u okviru projekta TITUS u obrađivanim materijalima još uvijek koriste razni sustavi pisanja koji još nisu uopće (npr. linearno pismo A) ili su nedostatno dešifrirani (poput keltiberskog pisma). Stoga je bilo potrebno uspostaviti jednoznačni i standardni kodni sustav za prikaz znakova, što komercijalni operacijski računalni sustavi do prije nekoliko godina nisu omogućavali.⁷⁸ Zato je i razvijen, kao što je navedeno, odgovarajući unikodni font.⁷⁹ Time su s današnjim razvojem sustava uglavnom bili riješeni problemi transliteracije odnosno prijenos izvornih grafičkih jedinica originalnog pisma u neutralne simbole (npr. brojeve) bez obzira na funkcionalne vrijednosti jedinica (grafematičke, fonološke i morfološke). Međutim, u slučaju dešifriranih pisama u znanstvenoj praksi korišteni transliteracijski sustavi temeljeni na latinici u stvarnosti podrazumijevaju fonetsku interpretaciju.

Za elektroničku pohranu tekstualnih podataka pri transliteraciji su stoga potrebni barem 16-bitni kodni sustavi kako bi se omogućila paralelna upotreba različitih transliteracija i istovremen prikaz originalnog i transliteriranog zapisa. Naime, nakon systemske analize pisma, tekst se može uređivati, čak i bez potpunog dekodiranja i fonetskog tumačenja izbornih znakova. Na taj način mogu se ostvariti metode kritičke usporedbe postojećih različitih kopija istoga teksta razvijenih u klasičnoj filologiji. Projekt TITUS kod pripreme takvih tekstova automatski prikazuje slike povezane s odgovarajućom transkripcijom teksta. Na taj način uvelike je olakšana fonološka, morfološka i sintaktička analiza povijesnih faza jezika, te omogućen lagan pristup sadržaju teksta radi filoloških, književno-povijesnih i povijesnih istraživanja.

⁷⁸Bunz, Carl Martin, Ibid.

⁷⁹Vidi bilješku 64.

Lingvistički sadržaj koji se nalazi u tekstovima pohranjen je u tekstualnoj bazi podataka projekta TITUS te je dostupan na dva temeljna načina.⁸⁰

Poziv na određeni dio teksta. Svi su tekstovi podijeljeni na više razina, tj. na poglavlja, odlomke, stihove, stranice (izdanja), retke i slično, najčešće u skladu s tiskanim izdanjem prema kojem je pripremljen tekst. Određenom dijelu teksta može se pristupiti preko elektroničkog formulara koji pokriva cijelu tekstualnu bazu podataka.

Upotreba programa za pretraživanje prema riječima, za koji postoje četiri načina, koji su svi slični. Važno je naglasiti da je u svim tekstovima svaka riječ povezana tako da se dvostrukim klikom na neku riječ pronalaze sve pojave tražene riječi (uključujući i one s manjim pravopisnim razlikama) i prikazuju konkordance pri čemu je ključna riječ prikazana u svojoj okolini, tj. kao KWIC (*keyword-in-context*) konkordance. Dobiveni rezultati mogu se ponovo koristiti kao početna točka za daljnje upite. Unutar sustava TITUS pretraživanje je moguće uz točno poklapanje izraza za pretraživanje s tekstom, zatim uz približno poklapanje, u kojem se pronalaze riječi s minimalnom razlikom u zapisu, te kroz tzv. nepotpuno (*fuzzy*– zamućen, zamagljen, neodređen) poklapanje, gdje se pri pretraživanju zadani elementi riječi smatraju samo strukturom, s bilo kojim nizom znakova dozvoljenim između zadanih znakova.

Zanimljivo je da od starih hrvatskih tekstova⁸¹ na tzv. starom hrvatskom (*Old Croatian*) postoji samo jedan tekst, i to „Ribanje i ribarsko prigovaranje“

⁸⁰Vidi <http://titus.fkidg1.uni-frankfurt.de/texte/etcs/slav/asloven/freisdg/freis.htm> (2. 9. 2014.).

⁸¹ Projekt TITUS, popis starih hrvatskih tekstova (Old Croatian): <http://titus.uni-frankfurt.de/indexe.htm?/unicode/unitest2.htm> ; <http://titus.uni-frankfurt.de/texte/texte2.htm#acroat> (3. 9. 2014.).

Petra Hektorovića iz 1568. godine, ali nije dostupan jer je „u pripremi“ (*under preparation*), unos podataka pripisuje se W. Vermeru iz Amsterdama. Istovremeno treba naglasiti da tekst toga spjeva postoji pripremljen još od 2002. godine u okviru projekta Hrvatska rječnička baština, te se njegova digitalna kopija u izvornoj grafiji nalazi u digitalnoj dokumentaciji toga projekta,⁸² dok se više različitih primjeraka digitalnog teksta toga spjeva može naći na nekoliko internetskih stranica, ali u transkribiranom i komentiranom obliku.⁸³

U nedostatku bilo kojeg drugoga teksta pisanog na hrvatskom jeziku iz ranijih stoljeća, kao primjer teksta sa slobodnim pristupom u okviru projekta TITUS prikazan je početak teksta Homerove Odiseje⁸⁴:

TITUS
Author: Hom.
Homer
Book: Od.
Odyssee

On the basis of the edition by
ArthurLudwich,
HomeriOdyssea,
vol. 1-2,
Lipsiae 1889-1891
electronically prepared by
Marina Benedetti,
Siena 2000;
TITUS version by J. Gippert,
Frankfurt a/M, 20.1.2001 / 29.10.2007

⁸²Digitalna dokumentacija projekta "Hrvatska rječnička baština i prikaz rječničkoga znanja", Zagreb, Filozofski fakultet Sveučilišta u Zagreb, Odsjek za informacijske i komunikacijske znanosti, Katedra za obradu prirodnoga jezika, leksikografiju i enciklopedistiku, Ivana Lučića 3, Zagreb, 2002.

⁸³ Projekt eLektire, <http://lektire.skole.hr/> (3. 9. 2014.); Stranice Odsjeka za informacijske i komunikacijske znanosti (uz dozvolu izdavača "Bulaja" kao nakladnika pretiska u digitalnom izdanju) <http://www.ffzg.unizg.hr/infoz/dzs/html/Hektor1.htm> (3. 9. 2014.); itd.

⁸⁴Projekt TITUS, tekst Odiseje. <http://titus.uni-frankfurt.de/texte/etcs/grie/homer/odyssee/odyss.htm> (3. 9. 2014.).

Verse: 1 Ἄνδράμοι ἐννεπε, Μοῦσα, πολύτροπον, ὃς μάλα πολλὰ
Verse: 2 πλάγχθη, ἐπεὶ Τροίης ἱερὸν πτολίεθρον ἔπερσε·
Verse: 3 πολλῶν δ' ἀνθρώπων ἵδεν ἄστεα καὶ νόον ἔγνω,
Verse: 4 πολλὰ δ' ὅγ' ἐπὶ νῆπιά θεν ἄλγεα δὲ κατάθυμόν,
Verse: 5 ἄρνυμένοσ' ἦν τε ψυχὴ καὶ νόστον ἐταίρων.
Verse: 6 ἀλλ' οὐδ' ὥς ἐτάρους ἐρρῦσα τοιέμενός περ·
Verse: 7 αὐτῶν γὰρ σφετέρῃσιν ἀτασθαλίῃσιν ὄλοντο,
Verse: 8 νήπιοι, οἳ κατὰ βοῦς Ὑπερίονος Ἥελίοιο
Verse: 9 ἦσθιον· αὐτὰρ ὁ τοῖσιν ἀφείλετο νόστιμον ἦμαρ.
Verse: 10 τῶν ἀμύθεν γε, θεά, θύγατερ Διός, εἰπέ καὶ ἡμῖν.

Kao zaključak, i ovdje je uočljivo, na gotovo jednak način kao i u projektu Gutenberg, uvodni dio teksta sadrži kompletnu referencu, s označenim načinom kodiranja te napomenom o transkriptoru.

Slog teksta korektno je grafički strukturiran izborom fonta te izborom znakova starogrčkog alfabeta sa svim dijakriticima i naglascima (oštri i meki hak, iota subscriptum, dijereza, sva tri naglaska, te oznakom poglavlja kao i posebnom oznakom broja svakoga stiha), tako da je vrlo pogodan za automatsko strukturiranje teksta.

Svaka riječ predstavlja poveznicu (*link*) kojom se otvara novi ekran s navođenjem gdje se sve (u svim jezicima) ta riječ pojavljuje.

Slike izvornog sloga tiskane knjige nisu dostupne.

4.2.3. Projekt *Perseus*

Projekt *Perseus*⁸⁵ (Perseus Digital Library⁸⁶), koji je također poznat i kao Perseus Hopper, projekt je digitalne biblioteke na Sveučilištu Tufts (Tufts

⁸⁵ Članak u Wikipediji, ThePerseus Project (Perseus Hopper).
http://en.wikipedia.org/wiki/Perseus_Project (2.9. 2014.).

⁸⁶ The Perseus Digital Library, <http://www.perseus.tufts.edu/hopper/> (2. 9. 2014.).

University⁸⁷). Projekt prikuplja digitalne zbirke iz humanističkog područja i smješten je na Odsjeku za klasičnu filologiju.

Projekt je utemeljen 1987. radi prikupljanja i prikaza građe potrebne za proučavanje i studije o staroj Grčkoj. Na internetu se pojavio 1995. godine, te je proširio svoj opseg. Današnje zbirke pokrivaju grčko-rimske klasike te englesku renesansu.

Djela na starogrčkom jeziku pohranjena su u beta-kodu (*beta code*⁸⁸), ali se mogu prikazivati i na više različitih standarda za transkripciju. Svi tekstovi za koje se smatra da ne podliježu autorskim pravima (*copyright*) dostupni su za slobodno preuzimanje na mreži u XML formatu. Neki od sadržaja ograničeni su pravima intelektualnog vlasništva.

Projekt Perzej je projekt s otvorenim izvornim kodom (*open-source*)⁸⁹ te pruža servise za interakciju s tekstualnim zbirkama. Iako u cjelini predstavlja integriranu okolinu za čitanje, pojedinačni servisi projektirani su modularno, te ih se može razvrstati u tri klase.

Kada je riječ o jezičnoj potpori, program je neovisan o jeziku, ali obuhvaća specifičnu podršku za grčki, latinski i arapski jezik. Za svaki konkretni tekst na bilo kojem od ta tri jezika (tekst u odgovarajućem kôdu ili tekst označen XML-om kompatibilnim s TEI standardom), sustav pruža usluge za automatsku lematizaciju (povezujući pojavnice s osnovnim oblicima riječi) i

⁸⁷Tufts University, privatno istraživačko sveučilište, Medford/Somerville, blizu Boston, Massachusetts, SAD. <http://www.tufts.edu/> (2. 9. 2014.).

⁸⁸ČlanakuikipedijiBetaCode, http://en.wikipedia.org/wiki/Beta_Code (2. 9. 2014.).

⁸⁹Prema: The Perseus Digital Library – Open Source <http://www.perseus.tufts.edu/hopper/opensource> (2. 9. 2014.).

morfološku analizu (prepoznajući, na primjer, da je latinska pojava *amor* imenica muškoga roda u nominativ jednine). Na široj razini, sustav također omogućuje korpusna istraživanja koja automatski daju frekvencijske podatke za pojavnice i leme cijeloga teksta koji obrađuje.

Sustav omogućuje i kontekstualizirano čitanje u kojem se mogu automatski uspostavljati relacije više razine koje proizlaze iz ugrađenoga znanja, te se mogu automatski ustanoviti relacije kojima se može odrediti je li, na primjer, neki tekst prijevod nekog drugog izvornog teksta.

Korisnici ne samo da mogu čitati odlomke iz tekstova, nego i koristiti paket alata za pretraživanje radi pretraživanja i pronalaženja tražene riječi, fraze ili teksta pisanog na bilo kojem od jezika koje sustav podržava. U pretraživanju se mogu uključiti i opcije koje omogućuju pretraživanje prema svim mogućim oblicima riječi za grčki, latinski ili arapski. Budući da ti jezici imaju bogatu fleksiju, taj servis predstavlja vrlo moćno sredstvo za obradu. Za klasične tekstove, koji imaju dobro standardiziranu citatnu shemu, korisnici se mogu kretati kroz tekst upisivanjem kanonske skraćenice (na primjer, Thuc. 1,24).

Programski kôd je proširljiv, te korisnici mogu uključiti svoje vlastite XML tekstove usklađene s TEI.

U sljedećem primjeru prikazan je jedan Bakhilidov⁹⁰ ditiramb pisan na starogrčkom jeziku. U lijevom je stupcu tekstualni prikaz prikazan u Unicodu, a u desnom stupcu je XML verzija prilagođena projektu Perseus.

<http://www.perseus.tufts.edu/hopper/text?doc=> <http://www.perseus.tufts.edu/hopper/xmlchunk?doc=Perseus%3At>

⁹⁰**Bakhilid**(grčki: Βακχυλίδης, oko 518. –450. pne.) bio jestarogrčki lirski pesnik, jedan od najistaknutijih pjesnika korske lirike. <http://sh.wikipedia.org/wiki/Bakhilid> (2.9. 2014.).

Ἀν|τηνορίδαι|ἡ|Ἑλένη|ς|Ἀπαίτησις

¹Ἀντήνορος] ἀντιθέου
!!!!!!] ρακοῖτις Ἀθάνας πρόσπολος,
×~~~~] Παλλάδος ὀρσιμάχου
~~~~~] χ] ρυσέας  
5~~~~~] ν' Ὀδυσσεῖ  
Λαρτιάδα Μενελ] άφ τ' Ἀτρεΐδα βασιλεῖ  
~~~~~] βαθύ] ζωνος Θεανῶ  
~~~~~] ον  
×~~~~~] προσήνεπεν  
10×~~~~~] ἐ] ὕκτιμέναν  
~~~~~]  
~~~~~] δων τυχόντες  
~~~~~] σὺν θεοῖς  
~~~~~] ος  
2~~~~~]  
~~~~~] μεσονύ] κτιος κέαρ  
3~~~~~]
⁴ἄγον, πατήρδ' εὐβουλος ἦρωσ
πάντα σάμαινεν Πριάμω βασιλεῖ
παίδεσσί τε μῦθον Ἀχαιῶν.
40 ἔνθα κάρυκες δι' εὐ-
ρεῖαν πόλιν ὀρνύμενοι
Τρώων ἀόλλιζον φάλαγγας
δεξιστρατονεῖς ἀγοράν.
παντᾷ δὲ διέδραμεν αὐδάεις λόγος·
45 θεοῖς δ' ἀνίσχοντες χέρας ἀθανάτοις
εὐχοντο παύσασθαι δυᾶν.
Μοῦσα, τίς πρῶτος λόγων ἄρχεν δικάϊων;
Πλεισθενίδας Μενέλαος γάρ νυῖ θελξιεπεῖ
φθέγξατ', εὐπέπλοισι κοινώσας Χάρισσιν·
50 ᾧ Τρῶες ἀρηϊφίλοι,
Ζεὺς ὑψιμέδων, ὃς ἅπαντα δέρκεται,
οὐ καί τις θνατοῖς μεγάλων ἀχέων,
ἀλλ' ἐν μέσφω κεῖται κιχεῖν
πᾶσιν ἀνθρώποις Δίκαν ἰθεῖαν, ἀγνᾶς
55 Εὐνομίας ἀκόλουθον καὶ πινυτᾶς Θέμτος·
ὀλβίων παιδὲς νιν αἰρεῦνται σύνοικον.
ἂδ' αἰόλοισι κέρδεσσι καὶ ἀφροσύναις
ἐξαισίους θάλλουσ' ἄθαμβῆς
ῥ' ὕβρις, ἅπλ[οῦτον] δύναιμι τεθοῶς
60 ἄλλοτριν ὥπασεν, αὐτὶς
δ' ἐς βαθὺν πέμπει φθόρον,
κεῖνα καὶ ὑπερφιάλους
Γᾶς παῖδας ὥλεσεν Γῆγαντας.

1 Column 30

2 7 lines are lost.

3 12 lines are lost.

4 Column 31

Bacchylides. The Poems and Fragments.
Cambridge University Press. 1905.

```

<TEI.2>
<text>
<body>
<div1 n="Dith" type="Book" org="uniform" sample="complete">
<div2 n="15" type="poem" org="uniform" sample="complete">
<head>
<title lang="en">Ode 15</title>
<title>Ἀν|τηνορίδαι|ἡ|Ἑλένη|ς|Ἀπαίτησις</title>
</head>
<note lang="en" resp="P" type="Papyr" anchored="yes" place="u
nspecified">Column 30</note>
<milestone n="str. 1" unit="strophe"/>
<l>Ἀντήνορος] ἀντιθέου</l>
<l>!!!!!!] ρακοῖτις Ἀθάνας πρόσπολος,</l>
<l>×~~~~] Παλλάδος ὀρσιμάχου</l>
<l>~~~~~] χ] ρυσέας</l>
<l n="5">~~~~~] ν' Ὀδυσσεῖ</l>
<l>Λαρτιάδα Μενελ] άφ τ' Ἀτρεΐδα βασιλεῖ</l>
<l>~~~~~] βαθύ] ζωνος Θεανῶ</l>
<milestone n="ant. 1" unit="strophe"/>
<l>~~~~~] ον</l>
<l>×~~~~~] προσήνεπεν</l>
<l n="10">×~~~~~] ἐ] ὕκτιμέναν</l>
<l>~~~~~]</l>
<l>~~~~~] δων τυχόντες</l>
<l>~~~~~] σὺν θεοῖς</l>
<l>~~~~~] ος</l>
<note lang="en" resp="P" type="Text" anchored="yes" place="un
specified">7 lines are lost.</note>
<milestone n="str. 2" unit="strophe"/>
<l n="22">~~~~~</l>
<l n="23">~~~~~] μεσονύ] κτιος κέαρ</l>
<note lang="en" resp="P" type="Text" anchored="yes" place="un
specified">12 lines are lost.</note>
<milestone n="ep. 2" unit="strophe"/>
<l n="36">~~~~~</l>
<note lang="en" resp="P" type="Papyr" anchored="yes" place="u
nspecified">Column 31</note>
<l>ἄγον, πατήρ δ' εὐβουλος ἦρωσ</l>
<l>πάντα σάμαινεν Πριάμω βασιλεῖ</l>
<l>παίδεσσί τε μῦθον Ἀχαιῶν.</l>
<l n="40">ἔνθα κάρυκες δι' εὐ-</l>
<l>ρεῖαν πόλιν ὀρνύμενοι</l>
<l>Τρώων ἀόλλιζον φάλαγγας</l>
<milestone n="str. 3" unit="strophe"/>
<l>δεξιστρατον εἰς ἀγοράν.</l>
<l>παντᾷ δὲ διέδραμεν αὐδάεις λόγος:</l>
<l n="45">θεοῖς δ' ἀνίσχοντες χέρας ἀθανάτοις</l>
<l>εὐχοντο παύσασθαι δυᾶν.</l>
<l>Μοῦσα, τίς πρῶτος λόγων ἄρχεν δικάϊων;</l>
<l>Πλεισθενίδας Μενέλαος γάρ νυῖ θελξιεπεῖ</l>
<l>φθέγξατ', εὐπέπλοισι κοινώσας Χάρισσιν.</l>
<milestone n="ant. 3" unit="strophe"/>
<l n="50">ᾧ Τρῶες ἀρηϊφίλοι,</l>
<l>Ζεὺς ὑψιμέδων, ὃς ἅπαντα δέρκεται,</l>
<l>οὐ καί τις θνατοῖς μεγάλων ἀχέων,</l>
<l>ἀλλ' ἐν μέσφω κεῖται κιχεῖν</l>
<l>πᾶσιν ἀνθρώποις Δίκαν ἰθεῖαν, ἀγνᾶς</l>
<l n="55">Εὐνομίας ἀκόλουθον καὶ πινυτᾶς Θέμτος:</l>
<l>ὀλβίων παιδὲς νιν αἰρεῦνται σύνοικον.</l>
<milestone n="ep. 3" unit="strophe"/>

```

The Annenberg CPB/Project provided support for entering this text.

Purchase a copy of this text (not necessarily the same edition) from [Amazon.com](http://www.amazon.com)

```
<l>ἀ δ' αἰόλοισ κέρδεσσι καὶ ἀφροσύναις</l>
<l>ἐξαισίους θάλλουσ' ἀθαμβήεις</l>
<l>Ὑβρις, ἃ πλ[οῦτον] δύνανται τε θοῶς</l>
<l n="60">ἀλλότριον ὤπασεν, αὐτὶς</l>
<l>δ' ἐς βαθὺν πέμπει φθόρον,</l>
<l>κεῖνα καὶ ὑπερφιάλους</l>
<l>Γαῖα παῖδας ὄλεσεν Γίγαντας.</l>
</div2>
</div1>
</body>
</text>
</TEI.2>
```

O samom projektu *Perseus* objavljen je velik broj autorskih članaka⁹¹ u kojima se opisuju ili analiziraju pojedini aspekti. Odjek projekta u pogledu načina pripreme podataka, lingvističkih i povijesnih podataka te pripreme povijesnih tekstova pokazuje da je riječ o projektu od posebnog značenja te da treba u što većoj mjeri koristiti iskustva i metode koje se koriste u projektu. I u pripremi podataka na projektu Hrvatska rječnička baština i hrvatski europski identitet, poželjno bi bilo da se iskustva koja se mogu odnositi i na taj projekt i primijene.

Stoga je, kao zaključak, i ovdje uočljivo da je u projektu *Perseus* na gotovo jednak način kao i u prethodno opisanim projektima u uvodni dio teksta dana kompletna referenca. Slog teksta korektno je grafički strukturiran izborom fonta te izborom znakova starogrčkog alfabeta sa svim dijakriticima i naglascima (oštri i meki hak, *iota subscriptum*, dijereza, sva tri naglaska, te oznakom poglavlja kao i posebnom oznakom broja svakoga stiha), tako da je vrlo pogodan za automatsko strukturiranje teksta.

⁹¹Samo na izvornoj stranici projekta objavljen je popis od 178 znanstvenih članaka s uputnicom na puni tekst pojedinog članka <http://www.perseus.tufts.edu/hopper/about/publications> (3. 9. 2014.).

XML verzija prilagođena je TEI standardu te ispravno strukturira tekst posebnim oznakama, tako da je sasvim pogodna za daljnju tekstualnu analizu te dodatno obilježavanje izravno u TEI.

Svaka riječ predstavlja poveznicu (*link*) kojom se otvara novi ekran s navođenjem gdje se sve (u svim jezicima) ta riječ pojavljuje.

Slike izvornog sloga tiskane knjige nisu dostupne.

4.3. Hrvatska iskustva

U Hrvatskoj postoji više internetskih stranica na kojima se može pronaći različita baštinska građa, međutim za potrebe ove disertacije dat ćemo prikaz triju najvažnijih karakterističnih stranica, odnosno portala dviju vodećih kulturnih ustanova, Hrvatskoga državnoga arhiva⁹² i Nacionalne i sveučilišne knjižnice,⁹³ te prikaz Portala Hrvatske rječničke baštine⁹⁴ koji je nastao kao rezultat znanstvenog projekta Hrvatska rječnička baština i prikaz rječničkoga znanja.⁹⁵

Prakse i iskustva ostvareni na navedenim portalima, koji se odnose na digitalizaciju hrvatske kulturne baštine, prikaz i pohranu digitalnih verzija izvornih dokumenata u slikovnom i tekstualnom obliku, te njihov prikaz na internetskim stranicama, mogu predstavljati značajan doprinos u analizi

⁹²Hrvatski državni arhiv, <http://www.arhiv.hr> (3. 9. 2014.).

⁹³Nacionalna i sveučilišna knjižnica u Zagrebu, <http://www.nsk.hr/> (3. 9. 2014.).

⁹⁴Portal hrvatske rječničke baštine, <http://crodiv.ffzg.hr/> (3. 9. 2014.).

⁹⁵Vidi bilješku 2.

postojećeg stanja te u odlučivanju o načelima prikaza, pohrane i označavanja baštinskih dokumenata, među kojima najvažniju građu predstavlja hrvatska rječnička građa sadržana u starijim hrvatskim rječnicima.

4.3.1. Arhivski informacijski sustav *ARHiNET* Hrvatskoga državnog arhiva

Hrvatski državni arhiv (HDA)⁹⁶, kao „središnja i matična arhivska ustanova, čuva, zaštićuje, stručno obrađuje i daje na korištenje pisanu baštinu nastalu radom središnjih tijela državne uprave, prosvjetnih, kulturnih, zdravstvenih i vojnih ustanova, te gradivo nastalo djelatnošću hrvatskog iseljništva i istaknutih pojedinaca i obitelji iz hrvatskog kulturnog kruga“.

Međutim, u pohranjenoj građi ima i dokumenata koji u užem smislu ne bi trebali biti predmet obrade ili pohrane prema programskim ciljevima Državnog arhiva kao takovog, ali su stjecajem okolnosti, raznim mehanizmima preuzimanja građe postali dio građe koja se čuva u arhivu. Među njima ima i neprocjenjivih rukopisa i dokumenata koji po svojoj biti ulaze u tzv. dokumente pisane na povijesnim jezicima, od kojih su mnogi dostupni i u digitalnom obliku.

Pristup svim tim digitaliziranim dokumentima dostupan je preko informacijskog sustava ARHiNET Hrvatskoga državnog arhiva⁹⁷ koji osim što predstavlja *online* registar arhivskih fondova i zbirki Republike

⁹⁶Hrvatski državni arhiv, <http://www.arhiv.hr/arhiv2/Onama/index.htm> (3. 9. 2014.).

⁹⁷Vidibilješku br. 5.

Hrvatske,⁹⁸ „omogućuje i pristup i pregledavanje digitalnih snimki najznačajnijih i najčešće korištenih dokumenata“. Također je dostupan njihovim transkriptima, a ponegdje i prijevodima, npr. s latinskoga na hrvatski jezik.

Kao primjer slikovnog prikaza jednog dokumenta pisanog na latinskom jeziku s vjernim transkriptom u tekstualni oblik i prijevodom na hrvatski, prikazan je rukopisni dokument koji sadrži oporuku kćeri Agape zadarskoga tribuna Dabra, datiranu 15. srpnja 999. u Zadru.⁹⁹

⁹⁸RegistararhivskihfondovaizbirkiRepublikeHrvatske, središnjanaacionalnaevidencijaarhivskogagrdivakojajesastavnodiosustavatrenutnosadržipodatkeo 15 300 fondovaizbirki, 100 000 drugiharhivskihjedinica (serija, predmetaidokumenata), 29 500 matičnihknjiga, 25 000 stvarateljai 7 500 imateljearhivskogagrdiva, 5 500 obavijesnihpomagalai 3 000 bibliografskihizvoraoarhivskomegradivudostupnihjavnostiputeminterneta. Tijekom 2010. objavljenesuidigitalnesnimkenajznačajnijihinajčešćekorištenihdokumenataidigitaliziranaobavijesnapomagalatako dajetrenutnoputemsustavajavnodostupnoiooko 120 000 digitalnihsnimakagrdivai 1 300 tekstualnihzapisa.ARHiNET Hrvatska, <http://www.arhiv.hr/arhiv2/ARHiNET/index.htm> (3. 9. 2014.).

⁹⁹ Monumenta antiquissima, HR-HDA-876-1, 999, 15. srpnja, U Zadru <http://arhinet.arhiv.hr/DigitalniArhiv/Monumenta/HR-HDA-876-1.htm> (3. 9. 2014.).

Slika 4.1 Oporuka kćeri Agape Zadarskoga tribuna Dabra 15. srpnja 999. u Zadru



Transkript

[In] Christi nom[ine] et eiusdem anno incarnationis DCCCCLXVIII, indictione XII, sub die fere quinto decimo mense iulii, in ciuitate Iadera, imperantibus piissimis et perpetuis augustis Constantino et Basilio, cathedra pontificali regente Basilio episcopo et domino Maio priore. Maturiori namque et salubri consilio a maioribus prouisum est, ut omnis homo ante dispensare et deliberare suas facultates debeat, priusquam repentine calamitatis casu periculo [irr]uente mortis metu urgeat, ne differente deliberatio aut articulum [lingue] loquendi careat ui aut distributor in ultima corporis angustia [sub celeri positus transi] tu optata sibiment testatione explicare non queat. [Idcirco ego tanti]lla Agape peccatrix, filia Dabro tribuni, diuino iubamine [freta, posita in infirmi]tate super proprium meum decumbens lectulum, licet uiribus corporis inualida, tamen mente deo auxiliante tuta, integro autem consilio hunc..... onis de hereditate nostra scribere rogauimus, ita demum ut post obitum meum ita maneant deo auctore firmum et stabilem.... ac si quidem pagina subscripta uidentur. In primis uolo et iubeo, [ut] cenaculo cum corte et quoquina et orto, que est post ipsa quoquina, sit in ecclesia sancti Chrisogoni; domum, ubi fornax fuit, cum orticello, que est ante ipsam,

Prijevod

U ime Kristovo i godine 999. od njegova utjelovljenja, indikcije XII., dana otprilike petnaestoga mjeseca jula, u gradu Zadru. Za vladanja vrlo pobožnih i vazda uzvišenih Konstantina i Bazilija, a dok je na biskupskoj stolici sjedio biskup Bazilije i gospodin Majus bio gradskim priorom. Zrelom naime i korisnom odlukom pobrinuše se predi, da svaki čovjek treba prije smrti da podijeli i odredi svoj imetak, prije nego što bude na to primoran slučajem nenadane nesreće i strahom pred smrću, da ne bi, dok odgađa odluku, jezik izgubio moć govora ili djelitelj, u krajnjoj tjelesnoj tjeskobi, pri brzom prelazu, ne bi mogao svjedočanstvom iskazati svoje posljednje želje. Radi toga je neznatna Agapa, grijehnica, kći tribuna Dabra, uzdajući se u božju pomoć, a shrvana bolešću, ležeći u svojoj vlastitoj postelji, iako slabe tjelesne snage, ipak božjom pomoću u punoj svijesti.....zamolismo, da se o našem nasljedstvu napiše ova oporuka tako.....da poslije moje smrti božjom pomoću ostane čvrsta i stalnakako se sve to vidi u dolje napisanom pismu. Najprije hoću i nalažem, da stan s oborom i kuhinjom i vrtom, koji se nalazi iza same kuhinje, pripadne crkvi svetog Krševana. Isto tako i kuća, gdje je bila peč, s vrtićem, koji je pred njom; isto tako i vrtić, koji je pred crkvom svetog

similiter; orticello, que est ante ecclesias sancti Chrisogoni, similiter, portiones de terras, que habeo in Uculo similiter; portiones de salinas, que habeo inter meos parentes, similiter; artatikio serico I, inuestitura de serico I, panno serico I, lisa I, pennulas II, pirestres XX, mappas III, antelectulos III, capilectulo I, mappula ad calicem operiendum I,....., racinas II, culcitras II, culcittrino I, capitales III, capitale de serico de serico I, caldarias maiores II, frixorias maiores II, cucuma..... mortario hereo I, sella ferrea I, vrceo hereo cum aquiminile, camastras II, deuterias II, archiscamno I, arcella I, buttes II, tina I, bussedo I, macinas pario I, turabulo hereo I, stagniolo I. Hec omnia sit in ecclesia sancti Chrisogoni, et sicut prelibata est, ita deo auctore permaneat, nullus umquam ea audeat uiolare. Quicumque uero nostre deliberationi contraire uoluerit, iratum habeat patrem et filium et spiritum sanctum et maledictionem a CCCX et VIII sancti patres et in inferno cum diabolo et angeli (!) eius ec Iuda Scariothen muneretur. Actum tempore, die et loco ac consulibus, ut supra dictum est. His palam testibus: [† Signum] manus Basilio, supradicto episcopo, [† Signum] manus Maius, prior supradicto. † Signum manus Dabro tribunus, frater eius. † Signum manus Andreas presbyter. † Signum manus Drago, tribunus de Spalato. † Signum manus Uiato. Tribunus de Apsaro. † Signum manus Maius tribunus Marra in Saco. † Signum manus Maius tribunus Selitula.

Izvor: Diplomatički zbornik Kraljevine Hrvatske, Dalmacije i Slavonije, sv.I, Zagreb, 1967, str. 48-49.
Priredio: Luka Vukušić, arhivist HDA

Krševana. Isto tako i dijelovi zemlje, koje imam u Ukulu. Dijelovi solila, koje imam kod svoje rodbine, isto tako. Jedan svileni pokrivač, jedna navlaka od svile, jedan komad svile, jedna lisičja koža, dva gunjca, dvadeset sudova, četiri stolnjaka, tri čilima pred krevetima, jedan jastuk, jedan rubac za pokrivanje vrča..... dvije krpe, dvije pernice, jedna perničnica, tri jastučnice, jedna jastučnica od svile.....od svile, dvije veće posude za grijanje soba, dvije veće zdjele za pečenje, jedan lonac za kuhanje.....jedna mjedena posuda za cvijeće, jedna stolica od željeza, mjeden vrč s rukomijom, dvije košare, dva teftera, jedan naslonjač, jedan kovčežić, dva suda za vino, jedan drveni sud, jedna trlica, jedan žrvanj, jedna posuda od kositera. Sve to neka bude crkvi svetoga Krševana i kako je gore spomenuto neka božjom pomoću i ostane. Nitko neka se nikada ne usudi da to povrijedi. A tko bi htio da se usprotivi našoj odluci, neka ga stigne gnjev oca i sina i svetoga duha i prokletstvo 318 otaca, te neka bude kažnjen u paklu s đavlom i njegovim anđelima i Judom Iskariotom. Učinjeno u doba, na dan i na mjestu pa u vrijeme poglavica, kako je gore rečeno. († znak) ruke Bazilija, gore spomenutog biskupa († znak) ruke Majus, gore spomenuti prior. † Znak ruke Dabro tribun, njegov brat. † Znak † ruke Andrija svećenik. † Znak † ruke Drago, tribun splitski. † Znak † ruke Vojat, tribun osorski. † Znak † ruke Majus tribun. † Znak † ruke Majus tribun.

Izvor: Nagy, Josip, Monumenta diplomatica I : Isprave iz doba hrvatske narodne dinastije, Zagreb, 1925. **Priredio:** Luka Vukušić, arhivist HDA

I ovdje kao zaključak možemo ustvrditi da je, kao i u prethodno opisanim projektima, u uvodnom dijelu teksta dana kompletna referenca. Slog teksta korektno je strukturiran izborom fonta tako da je bez daljnijega pogodan za automatsko strukturiranje teksta.

Tekst nije dostupan u označenoj XML verziji, te je dobar kandidat da ga se strukturira i označi prema načelima razrađenima u ovoj disertaciji.

Slika izvornog dokumenta dostupna je u odličnoj rezoluciji i kvaliteti.

4.3.2. Projekt *Digitalizirana baština* Nacionalne i sveučilišne knjižnice

Projekt *Digitalizirana baština* Nacionalne i sveučilišne knjižnice^{100,101} sadrži, kao što je već navedeno, digitalne zbirke zvuka prošlosti, digitaliziranu baštinu, stare hrvatske novine, stare hrvatske časopise, hrvatski arhiv *weba*, digitalni akademski repozitorij, te virtualnu zbirku djela Ruđera Boškovića. Ta zbirka, iako za sada malena i dobrim dijelom izvan konteksta ove disertacije, predstavlja značajan resurs na kojem će se moći primijeniti tehnike i tehnologije označavanja rječničkih i povijesnih tekstova razvijenih i prikazanih u ovoj disertaciji.

Pod pojmom digitalizacije, u navedenoj zbirci, prvenstveno se smatra digitalizirana slika, odnosno slika teksta u digitalnom formatu. Od dostupnih dokumenata odnosno knjiga za neke se navodi mrežna poveznica koja upućuje na digitalni tekst koji nije nužno nastao u Nacionalnoj i sveučilišnoj knjižnici, nego je rezultat nekih drugih projekata.

Kao primjer prikazujemo stranice Digitalizirane baštine koje se odnose na Marulićevu Juditu. Na osnovnoj poveznici koju predstavlja slika naslovnice, poveznica je na slike teksta koje se mogu prelistavati u visokoj rezoluciji. Posebna poveznica povezuje na tekst Judite objavljen na internetu („Na mreži:“) u izdanju naklade Bulaja, ali ta poveznica nije ispravna.

¹⁰⁰Vidi stranicu 11. i bilješke 6. i 7.

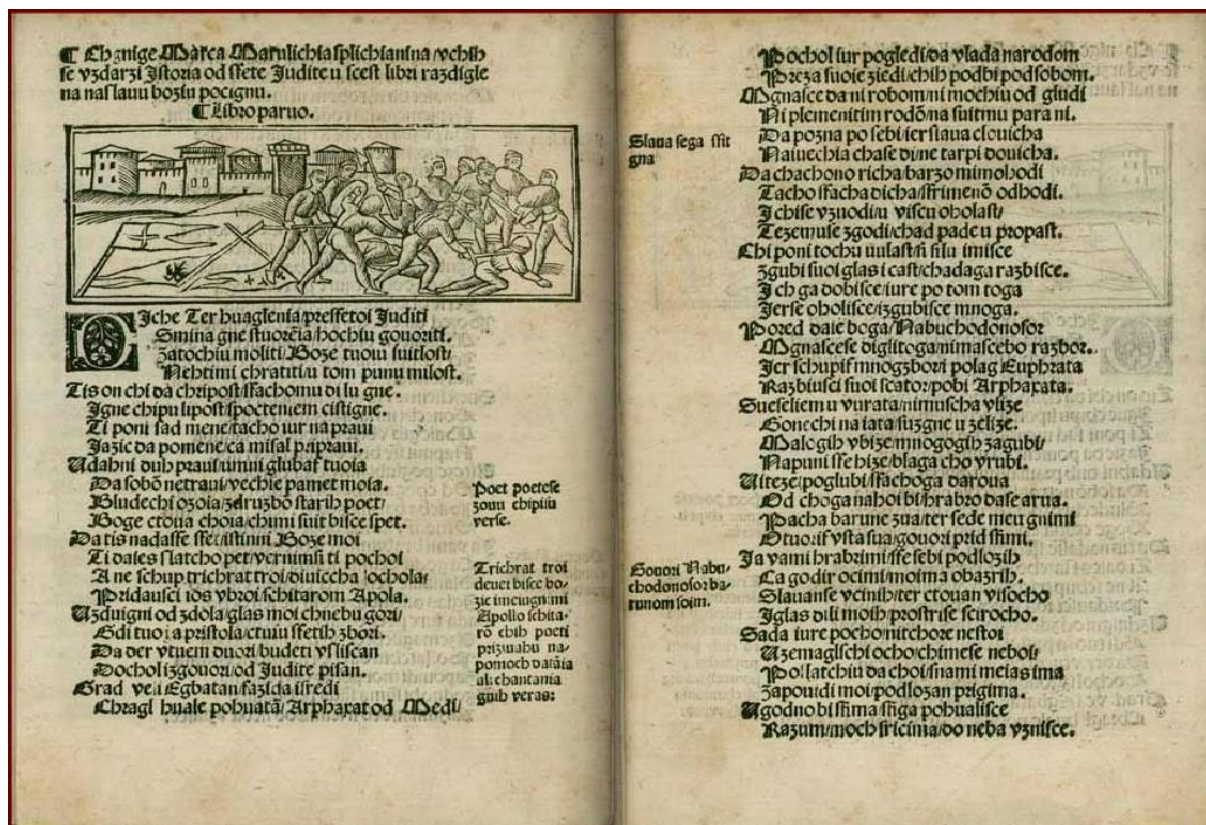
¹⁰¹Digitalizirana baština, Nacionalna i sveučilišna knjižnica, <http://db.nsk.hr/> (3.9. 2014.).

Slika 4.2 Digitalizirana baština Nacionalne i sveučilišne knjižnice poveznice na mrežna tekstualna izdanja Marulićeve Judite¹⁰²



¹⁰²NSK, Digitalizirana baština, <http://db.nsk.hr/HeritageDetails.aspx?id=873> (3. 9. 2014.).

Slika 4.3 Digitalizirana baština Nacionalne i sveučilišne knjižnice
slike digitalnog teksta Marulićeve Judite¹⁰³



Kao zaključak, i ovdje je, kao i u prethodno opisanim projektima, dostupna kompletna referenca.

Tekst uopće nije dostupan, ali postoje poveznice na nezavisne izdavače izvornoga teksta.

Slika izvornog dokumenta dostupna je u odličnoj rezoluciji i kvaliteti.

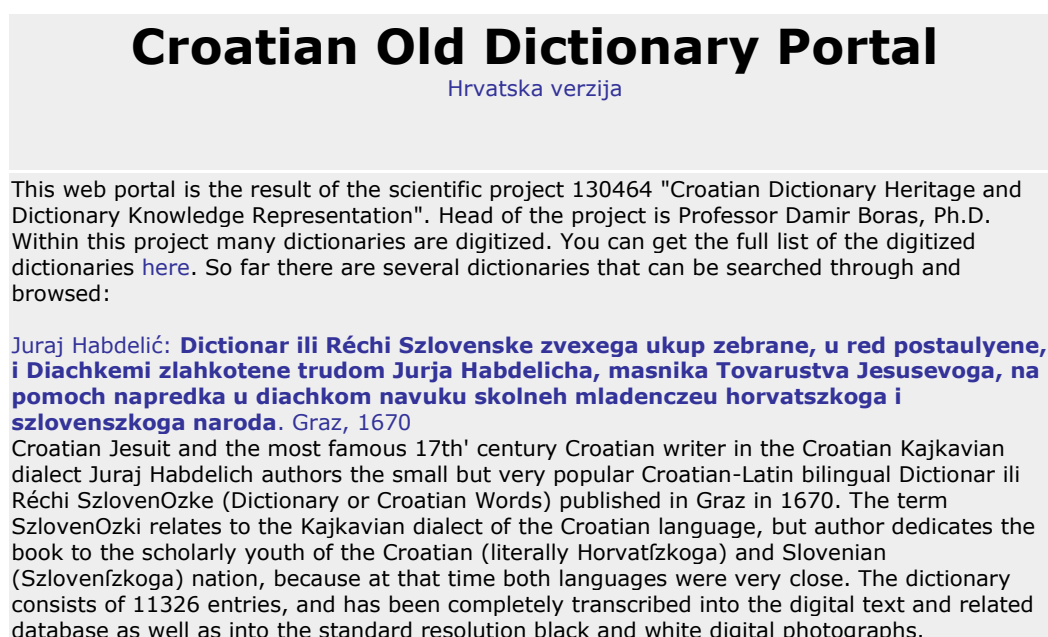
¹⁰³NSK, Digitalizirana baština, http://dk.nsk.hr/stara_knjiga/NSK_SK_ID03/ (3. 9. 2014.).

4.3.3. *Portal hrvatske rječničke baštine* projekta Hrvatska rječnička baština i prikaz rječničkoga znanja

Portal hrvatske rječničke baštine ogledni je portal projekta Hrvatska rječnička baština i prikaz rječničkoga znanja.¹⁰⁴ Portal ima i englesku verziju.¹⁰⁵ Na portalu je moguće pretraživati tekstove četiriju starih hrvatskih rječnika (vidi primjer) te pregledavati slike izdanja izvornoga rječničkog teksta.

Na portalu je omogućeno zasebno pretraživanje u svakom od dostupnih rječnika, ali i unakrsno pretraživanje. Po želji se pronađeni tekst prikazuje na dva načina: kao tekst sadržan u djelomično strukturiranoj bazi podataka ili kao slika odgovarajućih stranica rječnika.

Slika 4.4 Engleska verzija uvodne stranice Portala hrvatske rječničke baštine



¹⁰⁴Portal Hrvatske rječničke baštine, <http://cro dip.ffzg.hr/> (3.9. 2014.).

¹⁰⁵Croatian Old Dictionary portal, http://cro dip.ffzg.hr/default_e.aspx (3. 9. 2014.).

Bartol Kašić. **Manuscript without a title (Croatian (Čakavic) - Italian dictionary)**. Rome, 1599.

Although not undersigned, there is a strong scientific evidence that this manuscript was written in Rome, around 1600, by Croatian Jesuit Bartol Kašić who prepared this dictionary as a supplement for Croatian language learning to his Croatian grammar, published in 1604. (*Institutiones linguae illyricae*, Rome 1604.). Although it remained in the manuscript form, it was also scientifically proved that this dictionary was used as a basis for preparation of Croatian first bigger dictionary with the left Croatian side, the *Thesaurus of the Croatian language* or the Dictionary published by Italian Jacobi Micalia in Loretto and Ancona 1649-1651. The dictionary consists of 3613 entries and the digitized text is made searchable as well as browsable.

Libellus alphabeticus Cùm Nonnullis Catheticis addito vocabulario brevi latino, illyrico, germanico (Rerum communiter occurrentium - cum licentia superiorum.), s. l. (probably in Slavonia), 1756.

The digitized text of the dictionary is made searchable as well as browsable. The pictures of the original are also made browsable.

Faustus Verantius (Faust Vrancić). **Dictionarium quinque nobilissimarum Europae linguarum Latinae, Italicae, Germanicae, Dalmaticae et Ungaricae cum vocabulis Dalmaticis quae Ungari sibi usurparunt**. Venice, 1595.

The oldest Croatian printed dictionary published by famous Croatian author Faustus Verantius (Faust Vrancić) in Venice, 1595, is multilingual one covering Latin, Italian, German, Croatian (Dalmatian) and Hungarian with 5411 entries. The digitized text is made searchable and browsable. The pictures of the original are made browsable too.

Cross search through all the dictionaries

Zaključno, i ovdje je, kao i u prethodno opisanim projektima, dostupna kompletna referenca.

Tekst je dostupan u potpunosti, strukturiran je po jezicima te je vrlo pogodan za razna istraživanja, međutim ne postoji verzija za preuzimanje s mreže. Budući da je riječ o starijim hrvatskim rječnicima, na portalu nisu bili potrebni posebni prijevodi tekstova na hrvatski jezik.

Slika izvornog dokumenta dostupna je u odličnoj rezoluciji i kvaliteti.

Portal sadrži i popis od 18 rječnika digitaliziranih u okviru projekta pripremljenih u vrijeme prvobitnoga izdanja Portala.¹⁰⁶

¹⁰⁶Croatian Dictionary Heritage And Dictionary Knowledge Presentation, (Cdh) List Of Digitized Dictionaries And Bibliographies, <http://crodiip.ffzg.hr/rjecnici.pdf> (3. 9. 2014.).

4.4. Zaključak

Na temelju analiziranih primjera može se vidjeti da se u današnje vrijeme na internetskim portalima koji prikazuju baštinske dokumente na povijesnim jezicima uglavnom nastoje održati standardi načela prikaza sadržaja i znanja koji postoje u tim dokumentima.

Međutim, prikazi su, radi ograničenih tehničkih mogućnosti interneta u pogledu količine i brzine podataka, prilagođeni te su općenito smanjene rezolucije kad je riječ o digitalizaciji slike. Što se tiče upisa teksta i transkripcije, moguće je naći više različitih pristupa, ali se općenito može reći da je općeprihvaćeno načelo da se omogućuje dostup tekstu koji je dodatno obilježen nekim sustavom obilježavanja, najčešće u HTML ili XML formatu, ali se to sustavno ne provodi, tako da je najčešće pristup uvijek dostupan u najobičnijem tekstualnom formatu, a tek povremeno u nekom formatu za obilježavanje.

Razlog postojanju samo relativno malenog broja tekstova obilježenih u prikladnom formatu, s cjelovitom strukturom, leži u vremenski zahtjevnoj pripremi tako obilježenih tekstova, te relativno malom broju educiranih označitelja koji tekstove označuju, a posljedično i općenito skupoj pripremi.

Stoga je nužno razviti sustav koji će omogućiti automatsko obilježavanje tekstova na više razina, odnosno na razinama unosa izvornoga teksta, označavanje uočenih pogrešaka, transliteraciju, transkripciju, pretraživanje, prijevod na zajednički međujezik ili suvremeni jezik te napokon prikaz željene razine.

Problemi vezani uz razvijanje takvih sustava prikazani su u sljedećim poglavljima ovoga rada.

5. Obrada prirodnoga jezika (OPJ) i povijesni tekstovi

5.1. Uvod

Razvoj informacijsko-komunikacijske tehnologije (skraćeno ICT) omogućuje jednostavniji pristup i brži protok informacija, pohranu velikih količina podataka te sve jednostavnije rukovanje istima. Godine 2011. Hilbert i López¹⁰⁷ proveli su istraživanje u kojem su procijenili sposobnost tehnologije da pohrani, komunicira i obradi informacije, prateći 60 analognih i digitalnih tehnologija u razdoblju od 1986. do 2007. Prema istraživanju, sposobnost računala opće namjene da obradi podatke rasla je godišnje po stopi od 58 %. Također je na svjetskoj razini sposobnost dvosmjernih telekomunikacija godišnje rasla po stopi od 28 %, a pohrana informacija po stopi od 23 %.¹⁰⁸ Osim što je iz ovoga vidljivo da se sposobnosti informacijsko-komunikacijske tehnologije povećavaju, njena sve lakša primjena omogućuje prodiranje u sve segmente ljudskog djelovanja, pa tako i u kulturnu djelatnost. Stoga ne čudi činjenica da se sve više projekata koji uključuju kulturnu baštinu okreću upravo informacijsko-komunikacijskoj tehnologiji kako bi omogućili pristup, protok,

¹⁰⁷Hilbert, Martin; López, Priscila. 2011. "The World's Technological Capacity to Store, Communicate, and Compute Information." *Science* 332 (6025) (April 1): 60-65. doi:10.1126/science.1200970.

¹⁰⁸Hilbert i López 2011, Op. cit., Prema navedenom istraživanju, 2007. godine ljudska vrsta je mogla pohraniti $2,9 \times 10^{20}$ optimalno sažetih bajtova i komunicirati gotovo 2×10^{21} bajtova te izvršiti $6,4 \times 10^{18}$ naredbi u sekundi na računalima opće namjene.

pohranu i rukovanje baštinom. Prvi korak koji je potrebno učiniti jest digitalizacija baštine, odnosno pretvorba analognih podataka ili signala (npr. tiskani dokument) u digitalni, odnosno brojčani oblik,¹⁰⁹ kako bi se omogućila računalna obrada baštine.¹¹⁰

5.1.1. Digitalizacija kulturne baštine

U Hrvatskoj razlikujemo četiri osnovne baštinske djelatnosti: knjižničnu, arhivsku, muzejsku i konzervatorsko-restauratorsku. Njihov glavni cilj je prikupljanje, obrađivanje, istraživanje, pohranjivanje, zaštita i komuniciranje baštine.¹¹¹ Nekoliko je razloga zašto se provodi digitalizacija građe u arhivima, knjižnicama i muzejima, ali i u drugim ustanovama, na primjer na sveučilištima, fakultetima i institutima. Osnovni razlog pristupanju digitalizaciji građe je zaštita izvornika, koja može imati dva oblika. Kod aktivnog oblika digitalne preslike upotrebljavaju se umjesto samih izvornika kako bi se očuvali od mogućih oštećenja prilikom rukovanja. S druge strane kod pasivnog se oblika digitalne preslike koriste samo kao sigurnosne kopije koje mogu (barem djelomično) nadoknaditi gubitak u slučaju oštećenja ili uništenja izvornika. Sljedeći temeljni razlog digitalizacije građe je poboljšanje dostupnosti i mogućnosti korištenja građe, pogotovo ako se radi o rijetkim ili jedinstvenim objektima. Omogućavanjem mrežnog pristupa digitaliziranoj građi osigurava se

¹⁰⁹ Prema natuknici "Digitalizirati." *Hrvatski Jezični Portal*. Novi Liber. http://hjp.novi-liber.hr/index.php?show=search_by_id&id=f1dmXxQ%3D&keyword=digitalizirati (2.2. 2013.).

¹¹⁰ Prema natuknici "Digitize." *Oxford Dictionaries*. Oxford University Press. <http://oxforddictionaries.com/definition/english/digitize?q=digitise>. (2. 2. 2013.).

¹¹¹ Seiter-Šverko, Dunja; Križaj, Lana. "Digitalizacija kulturne baštine u Republici Hrvatskoj: od trenutne situacije prema nacionalnoj strategiji" NAPLE konferencija: Uloga narodnih knjižnica u nacionalnim politikama izgradnje društva znanja (Digitalizacija kulturne baštine i provedba multikulturalizma u knjižnicama), 5. 8. 2005. Supetar, Hrvatska. <http://www.min-kulture.hr/nadle/ind.htm>.

veća dostupnost kako stručnjacima tako i široj publici. Budući da pretvorba građe u digitalni oblik omogućava njezinu bržu i jednostavniju računalnu obradu, digitalizacijom se mogu stvoriti nove ponude i usluge korisnicima, koje su moguće jedino zahvaljujući novom mediju/okruženju u kojem se nalazi. Konačno, razlog pristupanju digitalizaciji građe može biti upotpunjavanje postojećeg fonda institucije, ako ta institucija iz različitih razloga procijeni da su joj određene digitalne preslike potrebne.¹¹²

Za digitalizaciju je prikladna različita vrste građe te je moguće digitalizirati tekstove, slike, zvuk, video, trodimenzionalne objekte, simulacije događaja itd. Međutim, za razliku od ostalih vrsta građe, i prema sadašnjem stanju tehnologije, samo se tekstovi mogu lako ili uopće pretraživati, dok se sustavi za pretragu ostalih vrsta građe tek razvijaju pa takvu građu nije lako ili nije uopće moguće pretraživati. Zbog tih se razloga digitalne preslike za takvu građu izrađuju uglavnom radi zaštite originala.

S druge se pak strane tekstovi mogu pretraživati, ali i automatski obrađivati. Upravo je to razlog zašto današnji projekti digitalizacije uglavnom teže pretvorbi tiskanih dokumenata u digitalne tekstove. Zahvaljujući sve lakšoj i širokoj upotrebi informacijsko-komunikacijske tehnologije u znanosti, ali i svakodnevnom životu, takvi projekti digitalizacije obuhvaćaju razna interdisciplinarna područja kao što su informacijske znanosti, društveno-humanistička informatika (engl. *digital humanities*, stariji naziv *humanities computing*), računalna lingvistika itd. Sva navedena područja uglavnom se oslanjaju na istraživanja digitalnih tekstova, neovisno o tome radi li se o

¹¹²Navedeno prema izvješću Radne grupe za digitalizaciju arhivske, knjižnične i muzejske građe Ministarstva kulture Republike Hrvatske 2006.

digitaliziranim tekstovima ili o tekstovima koji su izvorno nastali u digitalnom obliku.

5.1.2. Obrada prirodnoga jezika i povijesni tekstovi

Sve veća dostupnost povijesnih tekstova ima za posljedicu sve veće i veće zanimanje za primjenu metoda, a time i alata, za obradu prirodnoga jezika (engl. *natural language processing*) pri obradi povijesnih tekstova.

Ipak, povijesni tekstovi imaju nekoliko jezičnih značajki koje predstavljaju prepreke za primjenu alata obradu prirodnoga jezika u onom obliku u kakvom trenutno postoje. Pritom ćemo kao temeljnu klasifikaciju glavnih značajki tekstova preuzeti klasifikaciju koju je predložio autor Michael Piotrowski u svojoj knjizi „Obrada prirodnoga jezika za povijesne tekstove“.¹¹³

Prirodni jezik možemo definirati kao jezik kojim se govori i koji je nekome materinski.¹¹⁴ Naspram prirodnoga, postoje i formalni jezici koje oblikuju ljudi kako bi jednoznačno komunicirali s računalima. Budući da je danas većina informacija zapisana prirodnim jezikom, područje koje se bavi problemom obrade podataka kodiranih prirodnim jezikom naziva se obrada prirodnog jezika (engl. *natural language processing*), odnosno kraće OPJ (engl. *NLP*). Višeznačnost prirodnog jezika osnovna je razlika između prirodnih i formalnih jezika. Stoga je osnovni napor područja obrade prirodnog jezika

¹¹³Piotrowski, Michael. *Natural Language Processing for Historical Texts*. Edited by Graeme Hirst. Synthesis Lectures on Human Language Technologies 17. Morgan & Claypool Publishers, 2012.

¹¹⁴Premanatuknici “PrirodniJezik.”, *HrvatskiJezičniPortal.NoviLiber*. http://hjp.noviliber.hr/index.php?show=search_by_id&id=fVZvWhc%3D&keyword=jezik. (2. 2. 2014.).

formalizacija prirodnog jezika bez gubitaka informacije da bi on bio strojno obradiv.¹¹⁵

S druge strane, prema Piotrowskom¹¹⁶ pojam povijesnog teksta općenito možemo definirati kao tekstove koji su napisani povijesnim jezicima.

Problem se javlja kod definicije povijesnih jezika. Jedna od definicija „povijesti“ je *sveukupnost prošlih događaja*.¹¹⁷ Prema tome bi *povijesni* značio onaj koji pripada prošlim događajima, a tu možemo smjestiti jezike poput latinskog, starogrčkog i biblijskog hebrejskog jezik. Uzmemo li u obzir da su staroengleski, starovisokonjemački, starofrancuski i starohrvatski teško razumljivi ili potpuno nerazumljivi današnjim govornicima engleskoga, njemačkoga, francuskoga i hrvatskoga jezika, očito je da te jezike također možemo svrstati pod povijesne jezike. Međutim, suvremene inačice prethodno navedenih jezika nastajale su još od 14. stoljeća: novi engleski jezik nastaje sredinom 16. st.¹¹⁸; rani novovisokonjemački jezik pojavljuje se već u 14. st.¹¹⁹; novofrancuski seže u 17. stoljeće¹²⁰; novohrvatski potječe iz sredine 18. st.¹²¹. Za potrebe ovoga rada nije potrebno definirati povijesne etape pojedinih jezika, ali

¹¹⁵ **Ljubešić, Nikola.** “Pronalaženje događaja u višestrukim izvorima informacija.” Doktorska disertacija. Filozofski fakultet, Sveučilište u Zagrebu, 2009.

¹¹⁶ **Piotrowski**, op. cit. str. 1

¹¹⁷ **Vujić, Antun.** “Povijest.” *Proleksis Enciklopedija*. Leksikografski zavod “Miroslav Krleža”. <http://proleksis.lzmk.hr/42422/> (4. 2. 2013.).

¹¹⁸ Vujić, Antun, ed. 2013b. “Engleski Jezik.” *Proleksis Enciklopedija*. Leksikografski zavod “Miroslav Krleža”. <http://proleksis.lzmk.hr/55584/>. (4. 2. 2013.).

¹¹⁹ Vujić, Antun, ed. 2013c. “Njemački Jezik.” *Proleksis Enciklopedija*. Leksikografski zavod “Miroslav Krleža.”

¹²⁰ Vujić, Antun. 2013d. “Francuski Jezik.” *Proleksis Enciklopedija*. Leksikografski zavod “Miroslav Krleža”. <http://proleksis.lzmk.hr/667/>. (4. 2. 2013.).

¹²¹ Vujić, Antun. 2013e. “Hrvatski Jezik.” *Proleksis Enciklopedija*. Leksikografski zavod “Miroslav Krleža”. <http://proleksis.lzmk.hr/27217/> (4.2. 2013.).

je očito da, iako su djela jednog Shakespearea, Goethea, la Fontainea ili Šenoe relativno razumljiva današnjim govornicima pojedinih jezika, oni se od suvremenih tekstova mogu razlikovati po uporabi zastarjelica ili po stilu. Budući da se djela spomenutih autora razlikuju od suvremenih tekstova, primjena alata za obradu prirodnog jezika, čiji cilj su suvremeni jezici, neće biti u potpunosti prikladna u primjenama na stare tekstove. Prema tome, pojam povijesnih tekstova, a samim time i povijesnih jezika, možemo proširiti i izvan granica povijesne lingvistike, koja određuje granice početaka i krajeva povijesnih etapa pojedinih jezika.

5.2. Značajke suvremenih i povijesnih tekstova

Povijesni jezici su prirodni jezici, baš kao i jezici kojima danas govorimo i pišemo te se od današnjih načelno po ničemu ne razlikuju. To znači da će se tehnologija obrade prirodnoga jezika u primjeni na povijesne jezike vjerojatno suočiti s jednakim problemima na koje nailazi kada se obrađuju suvremeni tekstovi. U literaturi su često opisivani problemi razrješavanja homonimije¹²² i polisemije^{123,124,125}.

¹²²Homonimija je pojava da riječ ili više riječi istoga glasovnog sustava i istoga grafičkog oblika imaju različito značenje ili podrijetlo. Npr. "lak" može imati dva značenja: lak kao imenica u izrazu *lak za nokte* i lak kao pridjev koji opisuje predmet male težine ("Homonimija" 2013; "Lak (1)" 2013; "Lak (2)" 2013). Leksikografi različita značenja homonima uglavnom stavljaju kao posebne natuknice.

¹²³Polisemija znači mijenjanje riječi u vremenu, promjena značenja kao dijakronijska pojava u jeziku. Npr. "brod" može označiti određenu vrstu plovila na vodi. Osim toga, prije je označavalo mjesto na vodi, obično na rijeci gdje se može prijeći, pregaziti; gaz, plićak, prijelaz. Zastarjelo značenje možemo naći kao sastavni dio imena naseljenih mjesta, npr. Slavonski Brod ("Polisemija" 2013; "Brod" 2013). Leksikografi poliseme uglavnom stavljaju unutar iste natuknice.

¹²⁴ Piotrowski, Michael. 2012. *Natural Language Processing for Historical Texts*. Ed. Graeme Hirst. Synthesis Lectures on Human Language Technologies 17. Morgan & Claypool Publishers. <http://www.morganclaypool.com/doi/abs/10.2200/S00436ED1V01Y201207HLT017>. (str. 1).

Jezik se upotrebljava u različite svrhe i na različite načine, pa možemo govoriti o različitim područjima, žanrovima i stilovima. Danas se obrada prirodnoga jezika uglavnom obavlja nad mrežnim izvorima, često novinskim tekstovima dostupnima na internetu. Alati dizajnirani za novinske tekstove i na njima ućeni algoritmi imaju slabije rezultate kada se primijene na tekstove drugih domena ili žanrova.¹²⁶ Razlozi tome leže u razlikama na razini vokabulara, sintakse i semantike određenih područja i žanrova. Navedeni problemi mogu se preslikati i na cijelo područje obrade prirodnoga jezika, nevezano o tome radi li se o suvremenim ili povijesnim tekstovima.¹²⁷

Stoga će se, kako bi se dobio uvid u znaćajke obrade prirodnoga jezika na povijesnim tekstovima, u sljedećim poglavljima prvo opisati znaćajke

¹²⁵ *Word-sense disambiguation* // *Wikipedia, The Free Encyclopedia*. http://en.wikipedia.org/wiki/Word-sense_disambiguation (4.9. 2014.).

¹²⁶ Na primjer:

Bacchiani, M. et al. MAP adaptation of stochastic grammars. // *Computer Speech and Language*. 20, 1(2006), str. 41-68.

Daume III, H. Frustratingly easy domain adaptation. // *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Prag : Association for Computational Linguistics, 2007. Str. 256-263.

Gildea, D. Corpus variation and parser performance. // *Proceedings of EMNLP* / uredili L. Lee i D. Harman. Pittsburgh, 2001. Str. 167-202.

McClosky, D.; Charniak, E.; Johnson, M. Reranking and self-training for parser adaptation. // *Proceedings of COLING-ACL*. Sydney : Associationa for Computational Linguistics, 2006. Str. 337-344.

Sekine, S. The domain dependence of parsing. // *Proceedings od Applied Natural Language Processing (ANLP)*. 1997. Str. 96-102.

¹²⁷ Piotrowski, Michael. 2012. *Natural Language Processing for Historical Texts*. Ed. Graeme Hirst. Synthesis Lectures on Human Language Technologies 17. Morgan & Claypool Publishers. <http://www.morganclaypool.com/doi/abs/10.2200/S00436ED1V01Y201207HLT017>, (str. 2).

suvremenih jezika, koji su često ciljni jezici istraživanja i primjene obrade prirodnoga jezika, a zatim navedene značajke usporediti sa značajkama povijesnih jezika.

5.2.1. Značajke suvremenih tekstova

U svojoj knjizi Piotrowski¹²⁸ navodi četiri temeljne značajke suvremenih jezika na temelju kojih je moguće formalno uspoređivati status pojedinih jezika u smislu računalne obrade jezika:

1. standardna inačica
2. standardni pravopis
3. tekstovi u digitalnom obliku
4. postojeći resursi i alati za obradu prirodnog jezika.

5.2.1.1. Standardna inačica

Suvremeni jezici imaju svoju standardnu inačicu koja se upotrebljava u sredstvima informiranja. To su uglavnom jezici koji se poučavaju u osnovnim i srednjim školama i školama stranih jezika, a obilježava ih dobra dokumentiranost u rječnicima i gramatikama. Takva situacija omogućava jezikima razvoj alata i resursa za njihovu obradu te evaluaciju tih alata s obzirom na propisani standard.¹²⁹

¹²⁸Piotrowski, Michael, op. cit. str. 2.

¹²⁹Piotrowski, Michael, op. cit. str. 2.

5.2.1.2. *Standardni pravopis*

Suvremeni jezici imaju standardni pravopis kojeg se pridržava većina objavljenih tekstova. Autori tekstova mogu u rječnicima provjeriti kako se određena riječ ispravno piše. U većini slučajeva, alati za obradu prirodnoga jezika, koji se temelje na pravilima ili statistici, mogu računati na to da svaka leksička jedinica ima mali i konačan broj oblika koji se mogu pojaviti u tekstovima. Osim toga, standardizirani pravopis omogućava ispravak neispravno napisanih riječi.¹³⁰

5.2.1.3. *Tekstovi u digitalnom obliku*

Velika količina teksta na suvremenim jezicima dostupna je u digitalnom obliku i napisana standardnim pravopisom. Takvi tekstovi služe za razvoj alata i resursa za obradu prirodnoga jezika. Naravno, što je broj govornika pojedinog jezika veći, razmjerno će biti veća i količina teksta u digitalnom obliku u odnosu na druge jezike iza kojih stoje manje jezične zajednice.¹³¹ Možemo prema tome sa sigurnošću pretpostaviti da je na engleskom jeziku dostupna mnogo veća količina digitalnih tekstova nego na hrvatskom. Ipak, i za hrvatski jezik postoji količina novinskih i vladinih tekstova dostupna u digitalnom obliku dovoljna za početak razvoja alata i resursa za obradu jezika, a da prethodno nije potrebno posebno u tu svrhu digitalizirati tekstove.

5.2.1.4. *Postojeći resursi i alati za obradu prirodnog jezika*

Za takozvane „veće“ jezike, kao što su engleski, njemački, francuski ili španjolski, već postoje mnogi resursi i alati za obradu prirodnoga jezika: strojno čitljivi rječnici, korpusi, banke stabala, morfosintaktički označivači, parseri itd.,

¹³⁰Piotrowski, Michael, op. cit. str. 2.

¹³¹Piotrowski, Michael, op. cit. str. 2.

na kojima se može temeljiti daljnji razvoj. „Manji“ jezici uglavnom imaju bar neku vrstu resursa, makar to bio samo popis riječi za provjeru pravopisa.¹³²

5.2.2. Značajke povijesnih tekstova

Za povijesne jezike ne možemo reći da vrijede četiri navedene značajke. Poblje ćemo objasniti razloge na primjeru hrvatskoga jezika. Posebnu pozornost staviti ćemo na odsutnost standardiziranog pravopisa i što to točno znači za obradu prirodnoga jezika.

U svojem djelu Piotrowski¹³³ suprotstavlja četiri temeljne značajke povijesnih jezika temeljnim značajkama suvremenih jezika:

1. nedostatak standardne inačice
2. nedostatak standardnog pravopisa
3. nedostupnost tekstova u digitalnom obliku
4. nepostojeći resursi i alati za obradu prirodnog jezika za povijesne tekstove.

5.2.2.1. Nedostatak standardne inačice

Tek potkraj 19. i početkom 20. st., prateći i oponašajući događanja u vezi s nacionalnim jezicima u Europi, Hrvati su konačno uglavnom postigli suglasnost oko normiranja hrvatskoga književnoga jezika, u obliku kakav on uglavnom ima danas. Tada se ostvarila zamisao da se književni jezik temelji na jekavskome štokavskome, koji se ugledao na dubrovački jezik, odbacivši

¹³²Piotrowski, Michael, op. cit. str. 2.

¹³³ Prema Piotrowski, Michael. Op. cit. str. 3.

kajkavsko i čakavsko narječje kao standarde.¹³⁴

5.2.2.2. *Nedostatak standardnoga pravopisa*

U isto vrijeme kada se normirao hrvatski književni jezik, ustaljen je i današnji pravopis, koji je znatnije dotjerivan sredinom i koncem 20.st. Do tada nije bilo standardiziranog pravopisa, zbog čega se jedna riječ mogla pisati na više načina, ovisno o autorovim naklonostima. Na primjer, riječ „odreći“ ili „odreći se“ pisala se na razne načine u starim hrvatskim rječnicima. Jakov Mikalja je sredinom 17. stoljeća zapisuje u obliku „odreccife“¹³⁵, Andrija Jambrešić sredinom 18. stoljeća piše „odrechi“¹³⁶, Joakim Stulli početkom 19. st. „odrèchi“ i „odrèchi se“¹³⁷, a Mirko Divković i Dragutin Antun Parčić početkom 20. st. bilježe „odreći se“^{138,139}.

Povijesni tekstovi vrlo se razlikuju od suvremenih tekstova. Pravopis se u povijesnim tekstovima može razlikovati od dokumenata do dokumenta iz istoga razdoblja. Čak i isti autor često u istome tekstu piše jednu riječ na više

¹³⁴Vujić, Antun. 2013e. “Hrvatski Jezik.” *Proleksis Enciklopedija*. Leksikografski zavod “Miroslav Krleža”. <http://proleksis.lzmk.hr/27217/> (4. 2. 2013.).

¹³⁵Mikalja, Jakov; depropagandafide (Roma) Congregatio. 1649. *BlagoJezikaSlovinskogaIllySlovníkuKomuIzgorarajuse//Rjeci Slovinske Latinski, i Diacki. Thesaurus Linguae Illyricae Sive Dictionarium Illyricum in Quo Verba Illyrica Italice, & Latine Redduntur. Labore P. Jacobi Micalia: Societ. Jesu Collectum. Et Sumptibus Sacrae Congregationis De Propaganda Fide Impressum. Laureti: Apud Paulum, & Io: Baptistam Seraphinum.*

¹³⁶ Jambrešić, Andrija. 1742. *Lexicon Latinum Interpretatione Illyrica, Germanica, Et Hungarica Locuples, ... Digestum, Ab Andrea Jambressich.* Zagrabiae: Typis Academicis Societatis Jesu, per Adalbertum Wilh. Wesseli.

¹³⁷ Stulli, Joakim. 1806. *Rječosložje Iliričko-italijansko-latinsko.* Dubrovnik.

¹³⁸ Divković, Mirko. 1900. *Latinsko-hrvatski Rječnik Za Škole.* Izd. 2. U Zagrebu: Nakladom Kr. hrvatsko-slavonsko-dalmatinske zemaljske vlade.

¹³⁹ Parčić, Dragutin Antun. 1901. *Vocabolario Croato-italiano.* 3a ed. corretta ed aumentata. Zara: Tipografia editrice Narodni list.

načina. Nestandardizirani pravopis velika je prepreka obradi prirodnoga jezika jer se konvencionalne metode i tehnike oslanjaju na pretpostavku da su tekstovi na jednom jeziku napisani standardiziranim pravopisom.

Pitanja: „Kako sastaviti rječnik ako svaka riječ može biti napisana na više načina?“, „Što učiniti u slučaju kada su dvije različite riječi napisane na isti način?“, „Koju inačicu riječi odabrati da predstavlja sve ostale inačice iste riječi?“, predstavljaju statističkim metodama¹⁴⁰ značajne prepreke jer dolazi do raspršenosti podataka.

Stoga se, na primjer, u korpusu za učenje jedna riječ predstavlja inačicom A, iako su inačice B i C prema standardu ravnopravne i vjerojatnost pojavljivanja im je podjednaka. Iz navedenih razloga, statističke metode imaju problema s raspršenošću podataka s obzirom na to da jedna riječ može biti predstavljena s više inačica.

Osim toga, različite inačice riječi utječu na analizu teksta i metode obrade, kao što je npr. lematizacija, odnosno proces svođenja neke pojavnice na njezin temeljni oblik. Ako postoji više inačica jedne riječi, riječ će biti svedena na više lema (temeljnih oblika), umjesto da sve inačice budu svedene na jedan temeljni oblik.¹⁴¹

Dopuštene inačice u pravopisu postoje gotovo u svim suvremenim jezicima, ne samo u povijesnim. Na primjer, postoje britanski i američki

¹⁴⁰U području obrade prirodnoga jezika razlikujemo tradicionalne lingvističke metode temeljene na pravilima, novije statističke metode te hibridne metode koje kombiniraju prethodno navedene metode.

¹⁴¹Piotrowski, Michael. 2012. *Natural Language Processing for Historical Texts*. Ed. Graeme Hirst. Synthesis Lectures on Human Language Technologies 17. Morgan & Claypool Publishers. <http://www.morganclaypool.com/doi/abs/10.2200/S00436ED1V01Y201207HLT017>, (str. 3).

pravopisi (brit. engl. *colour* naspram amer. engl. *color*) ili njemački i švicarski pravopisi (njem. *die Straße* naspram švic. *die Strasse*). Međutim, različite inačice pisanja znatno su manje u suvremenim tekstovima nego u povijesnim. Osim toga, iz navedenih primjera vidljivo je da je zapravo riječ o posebnim standardima u Velikoj Britaniji i Sjedinjenim Američkim Državama, odnosno Njemačkoj i Švicarskoj, što znači da razlike u načinu pisanja ustvari podliježu pravilima, što ne predstavlja problem alatima za obradu prirodnoga jezika ako se prilagode određenoj inačici tog jezika.¹⁴²

Postoje, doduše, okolnosti kod suvremenih jezika gdje se pojavljuju različite inačice pisanja koje ne podliježu pravilima. To su pravopisne pogreške koje predstavljaju značajan problem pri pretraživanju teksta.¹⁴³

Osim toga, kod suvremenih jezika poseban problem predstavljaju nepoznate, iako ispravno napisane riječi, kao što su neologizmi. Iako mrežne tražilice danas mogu prihvatiti određeni broj inačica kako bi izašle na kraj s tipkarskim pogreškama, one i dalje pretpostavljaju da postoji samo jedan ispravan oblik svake riječi. Budući da takva pretpostavka nije prihvatljiva u obradi povijesnih tekstova, najveći je problem obrade povijesnih tekstova rukovanje pravopisnim inačicama.¹⁴⁴

Kao što je navedeno, jedna od značajki povijesnih tekstova jest mogućnost pojave iste riječi u tekstovima istoga autora napisane na više načina. U slučaju zbirke povijesnih dokumenata koje obuhvaćaju dulja razdoblja, problemi se povećavaju budući da se jezici i konvencije njihovoga zapisivanja

¹⁴²Piotrowski, Michael, op. cit. str. 3.

¹⁴³Piotrowski, Michael, op. cit. str. 3.

¹⁴⁴Piotrowski, Michael, op. cit. str. 3.

mijenjaju tijekom vremena.

Dodatni problem koji se javlja kod pretraživanja informacija na povijesnim tekstovima jest terminologija određenog vremena. U pretraživanju suvremenih tekstova (npr. mrežnih stranica) uglavnom se poznaju termini koji se pojavljuju u dokumentima koji se bave nama zanimljivim konceptima.¹⁴⁵ Na primjer, već spomenuta riječ „brod“¹⁴⁶ ima suvremeno i zastarjelo značenje. Kod pretraživanja mrežnih stranica, ako nas zanima brod kao plovilo, pretraživat ćemo prema riječi „brod“; ako nas zanima plićak odnosno gaz, pretraživat ćemo po riječima „plićak“ ili „gaz“.

Međutim, kod pretraživanja povijesnih tekstova situacija je slična onoj kada pretražujemo na stranom jeziku: uglavnom imamo ideju koje se riječi koriste, ali nismo sigurni. Uz to, u povijesnim dokumentima često susrećemo više od jednog jezika, na primjer latinski koji se kao *lingua franca* koristio nekoliko stoljeća.

Stoga je potrebno uočiti da, osim što povijesnim tekstovima nedostaje standardizirani pravopis, dodatne probleme predstavljaju pomaci u značenjima riječi (primjer riječi „brod“) te uporaba više od jednog jezika. Zbog toga bi bio koristan sustav pretraživanja povijesnih dokumenata koji bi omogućio uparivanje upita na suvremenom jeziku s povijesnim oblicima pronađenim u dokumentima. Na taj bi način korisnici mogli postavljati upite na suvremenom jeziku i pravopisu.

¹⁴⁵Piotrowski, Michael, op. cit. str. 3.

¹⁴⁶ “Brod.” 2013. Hrvatski Jezični Portal. Novi Liber. http://hjp.novi-liber.hr/index.php?show=search_by_id&id=f15kURY%3D&keyword=brod. (3. 2. 2013.).

5.2.2.3. Nedostupnost tekstova u digitalnom obliku

Osim što nije postojala standardizirana inačica hrvatskoga jezika i njegovoga pravopisa, dodatni problem obrade prirodnog jezika iz razdoblja prije standardizacije jest nedostupnost dovoljnih količina tekstova u digitalnom obliku, koji bi mogli poslužiti kao temelj za razvoj alata i resursa. Ako želimo stvoriti temeljne resurse (korpuse i alate za obradu), potrebno je prvo digitalizirati dokumente iz razdoblja koje nas interesira.

Treba, međutim, uočiti da će u budućnosti velika većina povijesnih dokumenata izvorno nastajati u digitalnom obliku. Može se, dapače, ustvrditi da danas praktički svi dokumenti nastaju izvorno u digitalnom obliku, iako se kasnije objavljuju i u analognom obliku (npr. na papiru, fotografiji, knjigama, novinama itd.). Hilbert i López¹⁴⁷ u svom su istraživanju procijenili koliko se informacija 2007. godine moglo pohraniti na analognim¹⁴⁸ i digitalnim¹⁴⁹ medijima za pohranu podataka. Prema njihovoj procijeni, 94 % pohranjenih podataka bilo je u digitalnom obliku, dok je papirnati oblik bio zastupljen sa samo 0,007 %.

5.2.2.4. Nepostojeći resursi i alati za obradu prirodnog jezika za povijesne tekstove

Ako u obzir uzmemo do sada navedeno, možemo ustvrditi da zapravo ne postoje osnovni alati i resursi za obradu povijesne inačice hrvatskoga jezika. Potrebno je stoga, za sastavljanje korpusa, prethodno digitalizirati tiskane

¹⁴⁷Hilbert, Martin, López, Priscila. 2011. "The World's Technological Capacity to Store, Communicate, and Compute Information." *Science* 332 (6025) (April 1): 60-65. doi:10.1126/science.1200970.

¹⁴⁸Primjeri analognih medija za pohranu podataka: video vrpca, fotografija, audiokaseta, negativ, knjige, tiskane novine itd.

¹⁴⁹Primjeri digitalnih medija za pohranu podataka: tvrdi disk, optička memorija, digitalna vrpca, prenosivi uređaji itd.

dokumente, zatim digitalnu presliku pretvoriti u tekst, koji će potom poslužiti kao temelj za razvoj osnovnih resursa za obradu prirodnoga jezika koji sadrži povijesne tekstove.

5.3. Zaključak

Povijesne tekstove karakteriziraju četiri temeljne značajke koje ih razlikuju od suvremenih tekstova. Prva je značajka da za jezike korištene u povijesnim tekstovima ne postoji standardna inačica jezika, za razliku od suvremenih tekstova u kojima upotrijebljeni jezici u velikoj većini imaju propisanu ili prihvaćenu standardnu inačicu. Druga je značajka da povijesnim tekstovima, naspram suvremenih tekstova, nedostaje propisani ili prihvaćeni standardni pravopis. Treća je značajka da je količina povijesnih tekstova dostupna u digitalnom obliku mnogo manja nego kad je riječ o suvremenim tekstovima. I konačno, kao četvrta značajka, odnosno kao posljedica prethodno navedenih značajki, nisu razvijeni resursi i alati za obradu prirodnog jezika koji bi omogućili obradu povijesnih tekstova. Stoga i postojeći resursi i alati čiji su algoritmi učeni na suvremenim tekstovima imaju slabije rezultate. Naime, primjena algoritama učenih na jednoj domeni daje slabije rezultate kada se isti primijene na tekstove drugih domena ili žanrova.¹⁵⁰

Stoga je potrebno razviti navedene alate za povijesne tekstove temeljene na metodama za obradu prirodnoga jezika, posebno tekstove na hrvatskom jeziku, na obradu kojih se i odnosi ovaj rad. Osim prethodno navedenih značajki povijesnih jezika, dodatne probleme predstavljaju pomaci u značenjima riječi te uporaba više od jednog jezika unutar jednog dokumenta. Također bi bilo

¹⁵⁰Vidi bilješku 81.

potrebno omogućiti uparivanje upita na suvremenom jeziku s povijesnim oblicima pronađenim u dokumentima, kako bi korisnici mogli postavljati upite na suvremenom jeziku i pravopisu. Sve su to problemi koje sustav temeljen na metodama za obradu prirodnog jezika na povijesnim tekstovima treba prevladati te će se na temelju ovih zaključaka predložiti i odgovarajući model.

6. Prijedlog modela

6.1. Uvod

U posljednjih nekoliko desetljeća u području obrade prirodnoga jezika razvijeno je mnogo jezičnih resursa, jezičnih alata i ostalih kompleksnijih sustava za obradu prirodnoga jezika. Međutim, razvoj se odvijao izolirano na projektima, sveučilištima i institutima, ali i od strane pojedinaca, gdje je nedostajalo komunikacije među članovima zajednice. Rezultat toga su resursi, alati i sustavi koji ne mogu međusobno razmjenjivati informacije.

Stoga se u najnovije vrijeme s jedne strane pokazala potreba za standardizacijom jezičnih tehnologija, ali i cijelih procesa razvoja. Interoperabilnost se javlja kao ključni pojam u ovoj fazi obrade prirodnoga jezika, čime se želi omogućiti međusobnu komunikaciju jezičnih podataka.

S druge strane, iz iskustva interoperabilnosti i suradnje uvidjelo se da automatske metode za opis prema nekoj zadanoj shemi imaju svoja ograničenja i domete te je jasno da nikad nije moguće razviti algoritam koji će potpuno točno označiti svaki tekst. Stoga je neophodan dodatni ljudski napor i vrijeme da se ti tekstovi dovedu do potpune točnosti (ili barem prihvatljive prema nekim statističkim pretpostavkama), kako bi bili prikladni za kasnije analize i obrade.

Stoga je u sljedećim poglavljima prikazano istraživanje koje je podijeljeno u dvije faze. U prvoj je fazi istraživanja predstavljen model prikaza znanja u rječnicima. Model prikaza znanja u rječnicima proveden je nad svih

sedam odabranih rječnika spomenutih u 3. poglavlju. U drugoj je fazi predložen model za automatsku odnosno poluautomatsku obradu tekstova, ispitana njegova točnost i testirano koliko automatska ili poluautomatska obrada ubrzava proces naknadnog označavanja, odnosno dovođenja teksta do prihvatljive razine točnosti označavanja. Za ovu fazu istraživanja odabran je della Bellin talijansko-latinsko-hrvatski rječnik iz 1785. godine jer je struktura natuknica unutar tog rječnika najsloženija. Svi elementi koji se mogu naći u natuknicama ostalih šest rječnika čine podskup elemenata natuknica della Bellinog rječnika. Stoga je za drugu fazu istraživanja odabran della Bellin rječnik zbog najsloženije strukture natuknica.

6.2. Model prikaza znanja u rječnicima

Kako bi se riješio problem nedostatka komunikacije između jezičnih podataka i postigla interoperabilnost jezičnih resursa, udružila su se dva velika međunarodna projekta. Jedan je projekt iz Sjedinjenih Američkih Država pod nazivom *Sustainable Interoperability for Language Technology* (SILT)¹⁵¹, čiji cilj je postojeću, ali fragmentiranu tehnologiju te resurse razvijene kao podrška jezičnoj tehnologiji pretvoriti u pristupačne, stabilne i interoperabilne resurse koji se u različitim područjima mogu lako ponovo koristiti.¹⁵² Drugi projekt dolazi iz Europe s nazivom *Fostering Language Resources Network* (FLaReNet)¹⁵³, te mu je cilj razviti zajedničku viziju u području jezičnih resursa

¹⁵¹Održiva interoperabilnost jezične tehnologije (prijevod autorice).

¹⁵²Sustainable Interoperability for Language Technology. <http://www.anc.org/LAPPS/SILT/index.html> (5. 9. 2014.).

¹⁵³Promicanje mreže jezičnih resursa (prijevod autorice).

i jezičnih tehnologija te promicati europsku strategiju za konsolidaciju sektora, time povećavajući konkurentnost na razini Europske unije i svijeta.¹⁵⁴ Cilj ove međunarodne suradnje jest uključiti članove zajednice iz područja jezičnih tehnologija, ali i ostale koji djeluju u srodnim područjima, kako bi surađivali. Suradnjom bi trebali stvoriti konsenzus vezan uz dijeljenje podataka i tehnologija za jezične resurse i aplikacije, raditi u smjeru interoperabilnosti postojećih podataka te, gdje to bude moguće, promicati standarde za označavanje i stvaranje resursa.¹⁵⁵

Jedno od glavnih postignuća navedene međunarodne suradnje jest ponuđena definicija pojma „interoperabilnost“ unutar područja jezičnih tehnologija.¹⁵⁶

Općenito se interoperabilnost može definirati kao stupanj do kojeg različiti sustavi, organizacije i/ili pojedinci mogu raditi zajedno kako bi postigli zajednički cilj. U računalnim sustavima interoperabilnost se uglavnom dijeli na dvije vrste: sintaktička interoperabilnost i semantička interoperabilnost. Kod sintaktičke interoperabilnosti svrha je omogućiti komunikaciju i razmjenu podataka putem određenih formata podataka, komunikacijskih protokola i slično. Važno je da dođe do razmjene informacija, ali nema jamstva da će tumačenje razmijenjenih informacija biti jednako. S druge strane, kod

¹⁵⁴Fostering Language Resources Network (FLaReNet). <https://www.flarenet.eu/> (5. 9. 2014.).

¹⁵⁵**Ide, Nancy; Pustejovsky, James; Calzolari, Nicoletta; Soria, Claudia.** The SILT and FLaReNet International Collaboration for Interoperability. Proceedings of the Third Linguistic Annotation Workshop, ACL-IJCNLP 2009, str. 178-181.

¹⁵⁶**Pustejovsky, James; Ide, Nancy.** “What Does Interoperability Mean, Anyway? Toward an Operational Definition of Interoperability,” 2010. <http://www.cs.vassar.edu/~ide/papers/ICGL10.pdf> (5. 9. 2014.).

semantičke interoperabilnosti sustavi imaju mogućnost automatski tumačiti razmijenjene informacije suvislo i točno kako bi dobili korisne rezultate, a tumačenje je moguće putem poštivanja zajedničkog referentnog modela za razmjenu informacija. Kod ove vrste interoperabilnosti bitno je da tumačenje razmijenjenih informacija bude jednako s obje strane komunikacije.¹⁵⁷

Kako bi se omogućila interoperabilnost jezičnih resursa, glavni smjer razvoja područja ide prema specificiranju apstraktnog modela podataka za strukturiranje jezičnih podataka, na koji se zatim mogu povezati sintaktička ostvarenja zajedno s povezivanjem na skup kategorija jezičnih podataka, koji komunicira sadržaj informacija odnosno jezični sadržaj. Stoga u kontekstu jezičnih resursa možemo sintaktičku interoperabilnost definirati kao mogućnost različitih sustava da obrade (odnosno čitaju) razmijenjene podatke bilo direktno bilo preko jednostavne konverzije. S druge strane, semantičku interoperabilnost možemo definirati kao mogućnost sustava da razmijenjene jezične informacije tumači suvislo te ih dosljedno upućuje na zajednički skup referentnih kategorija. U tom smislu jezični se resursi moraju više usredotočiti na semantičku nego na sintaktičku interoperabilnost.¹⁵⁸

Upravo su iz tog razloga u istraživanju korištene TEI oznake za označavanje strukture rječničkih natuknica. TEI (*Text Encoding Initiative*¹⁵⁹) konzorcij razvija i održava Smjernice (*Guidelines*) koje su namijenjene svima koji se bave pripremom i/ili obradom tekstualnih resursa u digitalnom obliku. Smjernice su ustvari preporuke konzorcija kako na odgovarajući način prikazati

¹⁵⁷Ibid.

¹⁵⁸Ibid.

¹⁵⁹Inicijativa za kodiranje teksta (prijevod autorice).

svojstva onih tekstualnih resursa koji se radi lakše računalne obrade moraju izričito označiti. Preporuke dolaze u obliku detaljnog opisa oznaka (*tag*) koje se unose u digitalni prikaz teksta, čime se obilježava struktura teksta i moguća druga svojstva koja se radi kasnije obrade žele eksplicitno označiti. Na taj se način implicitna tekstualna svojstva, u procesu koji se naziva kodiranjem (*encoding*), čine eksplicitnima. Kodiranje prema unaprijed definiranoj shemi omogućava razmjenu podataka bez gubitaka informacija i bez mogućnosti da će neka informacija biti krivo interpretirana. Budući da se TEI Smjernice temelje na XML-u, kao što se objašnjava u nastavku ovoga poglavlja, također su neovisne o računalnim programima i operativnim sustavima.¹⁶⁰

Smjernice se mogu primijeniti na tekstove svih prirodnih jezika, iz bilo kojega razdoblja i bilo koje književne vrste te bez ikakvog ograničenja na oblik i sadržaj. Također se mogu koristiti na kontinuirane tekstove bez razmaka, ali i na nekontinuirane tekstove, kao što su rječnici i korpusi. Upravo su zbog toga Smjernice odabrane kao temelj za označavanje strukture starih hrvatskih rječnika. Osim na digitalizirane tekstove, Smjernice se mogu primijeniti i na tekstove nastale u digitalnom formatu.¹⁶¹

Temelj Smjernica je *eXtensible Markup Language* (XML), jednostavan i fleksibilan jezik za označavanje koji se i inače koristi za razmjenu podataka na *webu* i drugdje.¹⁶² XML ima tri temeljne prednosti¹⁶³:

¹⁶⁰TEI Consortium (ur.). "iv. About These Guidelines".TEI P5: Guidelines for Electronic Text Encoding and Interchange. (Verzija 2.6.0). TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> (10. 7.2014.).

¹⁶¹Ibid.

¹⁶²eXtensible Markup Language (XML).<http://www.w3.org/XML/> (10.7.2014.).

1. stavlja naglasak na opisne oznake, a ne proceduralne
2. razlikuje koncepte sintaktičke točnosti i valjanosti s obzirom na upute za konkretan dokument
3. neovisan je o hardveru i softveru.

Uz to, XML omogućava prikaz i upis znakova prema standardu Unicode, što je nužno za starije hrvatske rječnike, budući da sadrže razne znakove i ligature koji danas ne pripadaju hrvatskoj abecedi, ali pripadaju tradiciji europske leksikografije. Navedenim standardom moguće je prikazati gotovo sve znakove koji se javljaju u starim hrvatskim rječnicima.

Sadržaj smjernica vrlo je širok, pa obuhvaća općenita poglavlja namijenjena svima koji se bave kodiranjem teksta, ali i specifična poglavlja za određene vrste tekstova te razne druge teme potrebne korisnicima. Općenita poglavlja bave se temama kao što su uvod u XML za potrebe korisnika smjernica, jezici i znakovi, infrastruktura TEI-a, zaglavlja u TEI-u, elementi dostupni na u svim dokumentima u TEI-u, te nestandardni znakovi i glifovi. U posebnim poglavljima smjernicama zasebno se opisuju: stihovi; izvedbeni tekstovi; transkripcija govora; rječnici; djela u rukopisu; prikaz primarnih izvora; kritičko izdanje; imena, datumi, ljudi i mjesta; tablice, formule, slikovni prikazi i bilježenje glazbe; jezični korpusi.¹⁶⁴

Za istraživanje provedeno u ovoj disertaciji najznačajnije je deveto poglavlje koje se bavi rječnicima. U poglavlju se definiraju oznake koje se mogu

¹⁶³TEI Consortium (ur.). “v. A Gentle Introduction to XML”. TEI P5: Guidelines for Electronic Text Encoding and Interchange. (Verzija 2.6.0). TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> (10. 7.2014.).

¹⁶⁴TEI Consortium (ur.). TEI P5: Guidelines for Electronic Text Encoding and Interchange. (Verzija 2.6.0). TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> (10. 7.2014.).

koristiti za označavanje rječničkih natuknica digitaliziranih rječnika, ali i drugih leksičkih resursa, kao što su, na primjer, glosariji.¹⁶⁵ Za označavanje rječničkih natuknica starih hrvatskih tiskanih rječnika korištene su upute iz tog poglavlja.

Pri obradi i označavanju rječnika treba se suočiti s dva glavna problema. Prvi je problem vezan uz strukturu natuknice, a drugi uz informaciju dostupnu u natuknici. Struktura rječničke natuknice razlikuje se od rječnika do rječnika, a najčešće i unutar jednog te istog rječnika. Kako bi shema za kodiranje mogla biti prikladna za različite strukture, jedini način je da se dozvoli da se bilo koji element može pojaviti bilo gdje u natuknici. Međutim, postoje i rječnici koji imaju dosljednu strukturu, te stoga shema mora podržavati i takve slučajeve.

Zbog tih je razloga u okviru TEI standarda bilo potrebno definirati dva različita elementa za rječničke natuknice: `entry` i `entryFree`. Elementom `entry` prikazuju se strukturirane natuknice, dok element `entryFree` sadrži iste komponente kao i element `entry`, ali se mogu slobodnije kombinirati. Preporuka Smjernica jest da se, gdje god je to moguće, prednost daje elementu `entry` naspram `entryFree`.¹⁶⁶ Za kodiranje natuknica starih hrvatskih rječnika upotrebljavat će se element `entry`.

Drugi problem koji pri označavanju rječničkoga teksta treba razriješiti jest način predstavljanja informacija u natuknici. Većina informacija u rječnicima je implicitna ili vrlo sažeta. Stoga je potrebno odabrati treba li prikazati informacije kao u izvorniku ili implicitne informacije učiniti eksplicitnima. Budući da su istraživačima oba načina jednako potrebna i

¹⁶⁵TEI Consortium (ur.). "9. Dictionaries". TEI P5: Guidelines for Electronic Text Encoding and Interchange. (Verzija 2.6.0). TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> (10. 7.2014.).

¹⁶⁶Ibid.

zanimljiva, shema mora podržavati i jedan i drugi način prikaza te također njihov međuodnos. Za neka je istraživanja važno zadržati izvorni prikaz natuknice, dok će za druga biti važne samo leksičke informacije što može za posljedicu imati promijenjen redoslijed pojavljivanja informacija, pri čemu je čak moguće i proširivanje, ali i izbacivanje određenih dijelova teksta. Moguća su istovremeno i oba pristupa, te je potrebno paralelno iskodirati tekst na oba načina i postaviti ih u međuodnos. Svi ti pristupi mogu se ostvariti upotrebom načela sadržanih u smjernicama TEI.¹⁶⁷

Za kodiranje tekstova natuknica starih hrvatskih rječnika najprikladniji je način da se zadrži izvorni redoslijed pojavljivanja informacija, dok će se dodatne informacije (koje se ne nalaze u izvornom tekstu) ukodirati u atributima elemenata. Na taj način razdvojiti će se izvorni sadržaj, koji se pojavljuje kao sadržaj elementa, od novog sadržaja odnosno značenja, koji će se dodati u atributu elementa.

6.2.1 Prikaz sheme

Na sljedećim primjerima prikazat ćemo strukturu rječnika i rječničkih natuknica pomoću TEI oznaka, te same oznake pobliže definirati i objasniti.

6.2.1.1 Zaglavlje

Svaki TEI dokument mora imati zaglavlje u kojem se mogu bilježiti osnovni podaci o samom kodiranom tekstu, izvori na kojima se temelji, principi kodiranja te voditi evidencija revizija učinjenih u dokumentu. Obavezan element koji zaglavlje mora sadržavati jest `fileDesc`, koji sadržava bibliografski opis

¹⁶⁷Ibid.

TEI dokumenta.¹⁶⁸ Na primjeru Mažuranić-Užarevićevog rječnika prikazat ćemo osnovne elemente zaglavlja te ih pobliže objasniti.

Korijenski element za zaglavlje jest `<teiHeader>`. Unutar svakog elementa može se definirati kojim je jezikom pisan. Stoga ćemo definirati da je zaglavlje pisano hrvatskim jezikom tako da ubacimo atribut `xml:lang="hr"` u element `<teiHeader>`. Obavezan element zaglavlja jest `<fileDesc>`, koji predstavlja bibliografski zapis digitalnog teksta. Sve je oznake potrebno zatvoriti, kao što se vidi na sljedećem primjeru:

```
<teiHeader xml:lang="hr">  
  
<fileDesc>  
  
</fileDesc>  
  
</teiHeader>
```

Jedini obavezni elementi unutar elementa `<fileDesc>` jesu `<titleStmt>`, `<publicationStmt>` i `<sourceDesc>` koje ćemo u nastavku opisati.

Element `<titleStmt>` sadrži informacije vezane za naslov. Osim samog naslova, mogu se zabilježiti i osobe ili institucije odgovorne za sadržaj elektroničkog dokumenta. Element `<title>` bilježi naslov, dok nekoliko drugih elemenata mogu bilježiti odgovorne osobe ili institucije. Prethodnom analizom strukture svih rječnika ustanovljeno je da treba zabilježiti autore sadržaja odnosno rječnika (`<author>`), urednike TEI dokumenta (`<editor>`)

¹⁶⁸TEI Consortium (ur.). "2. The TEI Header". TEI P5: Guidelines for Electronic Text Encoding and Interchange. (Verzija 2.6.0). TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> (10. 7.2014.).

te glavnog istraživača projekta (<principal>). Unutar svakog elementa, elementom (<persName>) je istaknuto da je riječ o imenima pojedinaca, jer je moguće obilježiti i institucije. Unutar imena eksplicitno su označena osobna imena (<forename>), prezimena (<surname>) i titule (<roleName>) pojedinaca. Navedena struktura elemenata i njihov sadržaj izgledaju ovako:

```
<titleStmt>
```

```
  <title    xml:lang="hr">Njemačko-ilirski    slovar:  
digitalno izdanje</title>
```

```
  <title          xml:lang="de">Deutsch-ilirisches  
Wörterbuch: digitale Ausgabe</title>
```

```
  <title  xml:lang="en">German-Croatian dictionary:  
digital edition</title>
```

```
<author>
```

```
<persName>
```

```
<forename>Ivan</forename>
```

```
<surname>Mažuranić</surname>
```

```
</persName>
```

```
</author>
```

```
<author>
```

```
<persName>
```

```
<forename>Jakov</forename>
```

<surname>Užarević</surname>

</persName>

</author>

<editor>

<persName>

<roleName>prof. dr. sc.</roleName>

<forename>Damir</forename>

<surname>Boras</surname>

</persName>

</editor>

<editor>

<persName>

<roleName>doc. dr. sc.</roleName>

<forename>Nikola</forename>

<surname>Ljubešić</surname>

</persName>

</editor>

<editor>

<persName>

```

<forename>Petra</forename>

<surname>Bago</surname>

</persName>

</editor>

<principal>

<persName>

<roleName>prof. dr. sc.</roleName>

<forename>Damir</forename>

<surname>Boras</surname>

</persName>

</principal>

</titleStmt>

```

Sljedeći obavezni element je <publicationStmt> u kojem se nalaze informacije vezane uz objavljivanje ili distribuciju elektroničkog teksta odnosno TEI dokumenta. Odabrani elementi unutar navedenog elementa su izdavač (<publisher>), adresa izdavača (<address>), mjesto izdavanja (<pubPlace>) te datum izdavanja (<date>). Navedena struktura elemenata i njihov sadržaj izgledaju ovako:

```

<publicationStmt>

```

<publisher>Filozofski fakultet Sveučilišta u Zagrebu</publisher>

<address>

<addrLine>Filozofski fakultet Sveučilišta u Zagrebu</addrLine>

<addrLine>Ivana Lučića 3</addrLine>

<addrLine>10000 Zagreb</addrLine>

<addrLine>Hrvatska</addrLine>

</address>

<address xml:lang="en">

<addrLine>Faculty of Humanities and Social Sciences</addrLine>

<addrLine>Ivana Lučića 3</addrLine>

<addrLine>10000 Zagreb</addrLine>

<addrLine>Croatia</addrLine>

</address>

<pubPlace>Zagreb</pubPlace>

<date when="2014">2014</date>

</publicationStmt>

Sljedeći element korišten u ovom dijelu zaglavlja jest <notesStmt>, neobavezni element unutar kojeg se slobodnim tekstom mogu nalaziti dodatne informacije. Slobodni tekst mora se nalaziti unutar elementa <note>. Unutar sadržaja elementa upotrijebili smo element <persName>, kako bi eksplicitno označili spomenute osobe. Navedena struktura elemenata i sadržaj vezan uz projekt unutar kojeg je izrađen dokument izgledaju ovako:

```
<notesStmt>
```

```
<note>Ovaj dokument pripremljen je u okviru
znanstvenog projekta "Hrvatska rječnička baština i
hrvatski europski identitet" (130-1301679-1380)
glavnog istraživača <persName><roleName>prof. dr.
sc.</roleName><forename>Damira</forename><surname>Bor
asa</surname></persName>. Financiran je od strane
Ministarstva znanosti, obrazovanja i sporta Republike
Hrvatske <date from="2007" to="2013">od 2007. do
2013. godine.</date></note>
```

```
<note xml:lang="en">This document is prepared as
part of the "Croatian dictionary heritage and
Croatian European identity" scientific project (130-
1301679-1380) of the main researcher
<persName><roleName>prof.</roleName><forename>Damir</
forename><surname>Boras</surname>,
<roleName>PhD</roleName></persName>. This project is
funded by the Croatian Ministry of Science and
Technology <date from="2007" to="2013">from 2007.
until 2013.</date></note>
```

</notesStmt>

Sljedeći obavezni element <sourceDesc> sadrži informacije izvornog teksta na kojem se temelji digitalni dokument. Budući da je obavezan element, a TEI oznake su namijenjene ne samo digitaliziranim tekstovima već i tekstovima koji nastaju u elektroničkom obliku, u ovom je slučaju potrebno dodati tekst, na primjer „Izvorno elektronički tekst“. Elementom <bibl> moguće je obilježiti bibliografske informacije izvora. Od već navedenih elemenata u ovom elementu mogu se koristiti i elementi <title>, <author>, <publisher>, <pubPlace> te <date>.

U naslovu je moguće atributom `type` odrediti radi li se o glavnom naslovu ili podnaslovu. Osim toga, dodan je element <choice> kojim se mogu prikazati alternativni načini kodiranja određenog dijela teksta. Tako se unutar njega nalazi element <orig> kojim se označava izvorni zapis teksta, dok se elementom <reg> označava regularni zapis istog dijela teksta, u ovom slučaju današnju grafiju određenog znaka. Na primjer, u naslovu Mažuranić-Užarevićevog rječnika napisana je riječ `Deutfch`, koja se danas piše `Deutsch`. Stoga je eksplicitno označena izvorna grafija, nakon koje slijedi današnja grafija:

<choice>

<orig>Deutfch</orig>

<reg>Deutsch</reg>

</choice>

Navedena struktura elementa <title> i sadržaj vezan uz rječnik je sljedeći:

```
<title type="main" xml:lang="de">

<choice>

<orig>Deutfch</orig>

<reg>Deutsch</reg>

</choice>-

<choice>

<orig>ilirifches</orig>

<reg>ilirisches</reg>

</choice>

Wörterbuch

</title>
```

Osim mogućnosti paralelnog zapisa izvornog i suvremenog teksta, elementom <choice> moguće je zabilježiti gdje se u tekstu pojavljuju kratice (<abbr>) te njihove pune nazive (<expn>). Na primjer, u nazivu izdavača nalazi se kratica k. za termin königlichen, odnosno kraljevski, a oznaka izgleda ovako:


```
<choice>

<abbr>k.</abbr>

<expan>königlichen</expan>

</choice>
```

Treća mogućnost navedenog elementa je eksplicitno zabilježavanje pogreške u tekstu (<sic>) te njezinog ispravka (<corr>). Tako u Vrančičevom rječniku možemo naći pogrešku u naslovu gdje je umjesto Dalmaticæ napisano Dalmatiæ, a primjer strukture i sadržaja elementa jest sljedeći:

```
<choice>

<sic>Dalmatiæ</sic>

<corr>Dalmaticæ</corr>

</choice>
```

Potpuna struktura i sadržaj elementa <sourceDesc>zabilježeni su na ovaj način:

```
<sourceDesc>

<bibl>

<title type="main" xml:lang="de">

<choice>

<orig>Deutfch</orig>
```

```
<reg>Deutsch</reg>

</choice>-

<choice>

<orig>ilirifches</orig>

<reg>ilirisches</reg>

</choice>

<choice>

<orig>Wörterbuch</orig>

<reg>Woerterbuch</reg>

</choice>

</title>

<title type="main" xml:lang="hr">

<choice>

<orig>NĚMAČKO</orig>

<reg>NJEMAČKO</reg>

</choice>-ILIRSKI SLOVAR

</title>

<author>

<persName>
```

<forename>Ivan</forename>

<surname>Mažuranić</surname>

</persName>

</author>

<author>

<persName>

<forename>Jakov</forename>

<surname>Užarević</surname>

</persName>

</author>

<publisher xml:lang="de">Verlag und Druck der

<choice>

<abbr>k.</abbr>

<expn>königlichen</expn>

</choice>

<choice>

<abbr>priv.</abbr>

<expn>privilegierten</expn>

</choice>

<choice>

<orig>ilirifchen</orig>

<reg>ilirischen</reg>

</choice>

National-Buchdruckerei von

<persName>

<roleName>Dr.</roleName>

<forename>Ljudewit</forename>

<surname>Gaj</surname>

</persName>

</publisher>

<publisher xml:lang="hr">Naklada i tisak
kraljevske privilegirane ilirske nacionalne
knjigotiskare

<persName>

<roleName>Dra.</roleName>

<forename>Ljudevita</forename>

<surname>Gaja</surname>

</persName>

</publisher>

<pubPlace xml:lang="de">Agram</pubPlace>

<pubPlace xml:lang="hr">Zagreb</pubPlace>

<date when="1842">1842</date>

</bibl>

</sourceDesc>

Posljednji element u zaglavlju je element <encodingDesc>, neobavezan element koji se nalazi u zaglavlju svih označenih rječnika projekta „Hrvatska rječnička baština i hrvatski europski identitet“. Inače se u njemu opisuje odnos između digitaliziranog teksta i njegovog izvora. U ovom se istraživanju koristi za opis projekta (<projectDesc>) u sklopu kojeg su označavani rječnici:

<encodingDesc>

<projectDesc>

<p>Od <date when="1595">1595. g.</date> kad je objavljen prvi tiskani hrvatski rječnik <persName><forename>Fausta</forename><surname>Vrančić</surname></persName> pa <date notBefore="1900" notAfter="1925">do početka 20. st.</date> u Hrvatskoj je objavljeno preko 300 različitih jednojezičnih, dvojezičnih, višejezičnih, enciklopedijskih, tehničkih, vojnih i drugih rječnika koji predstavljaju neprocjenjivu hrvatsku rječničku

baštinu od kojih je većina nedostupna za javnu upotrebu. Upravo su stari hrvatski rječnici središte ovog znanstvenog projekta pod nazivom "Hrvatska rječnička baština i hrvatski europski identitet" (130-1301679-1380). Samo je ograničen broj tih rječnika strojno obrađen u znanstvenom projektu 130464 "Hrvatska rječnička baština i prikaz rječničkoga znanja" (<date from="2002" to="2005">2002-2005</date>) na koji se ovaj projekt nastavlja i čiji su rezultati djelomično objavljeni na "Portal hrvatske rječničke baštine". Oba projekta financiralo je Ministarstvo znanost, obrazovanja i sporta Republike Hrvatske, a glavni istraživač bio je <persName><roleName>prof. dr.</roleName><forename>Damir</forename><surname>Bora s</surname></persName>.</p>

<p>Nastavljajući se na dotadašnji rad, cilj projekta bio je dodatno istražiti hrvatsku rječničku baštinu proširivši istraživanje <date notAfter="1945">do 1945. g.</date> Taj iznenađujuće velik broj rječnika objavljen u velikom broju europskih gradova (u prvim stoljećima uglavnom izvan hrvatskih prostora) potvrđuje pretpostavku da su Hrvatska i hrvatski jezik (u svim tadašnjim nazivima i idiomima) bili prisutni u Europi, da su bili neodvojivi dio europskog identiteta te da su svojim

sadržajem povratno utjecali na svijest o hrvatskom identitetu u Europi.</p>

<p>Cilj je projekta bio i dodatno ispitati recepciju tih rječnika u nas i u Europi, utvrditi razloge njihova objavljivanja, dopuniti izrađenu bibliografiju te dalje popuniti postojeću digitalnu bazu slika i tekstova najvažnijih hrvatskih starih dvojezičnih i višejezičnih rječnika te razviti računalne metode za dodatnu rječničku, jezičnu, gramatičku, terminološku analizu rječničkih tekstova, kao i usporednu i temporalnu analizu između pojedinih rječnika. Cilj je bio i istražiti strukturu rječničkoga znanja sadržanoga u njima s posebnim naglaskom na stručnu terminologiju te razviti strojne sustave koji bi omogućili takvu analizu za hrvatski jezik. Na znanstveno utemeljenim načelima se prezentira tekstualna i slikovna baza izvornih hrvatskih rječničkih tekstova i njihov europski identitet.</p>

<p>Projekt je za Hrvatsku važan u znanstvenom, kulturnom i političkom kontekstu jer je pokazao da je Hrvatska u području leksikografije i grafičke rječničke prezentacije bila ravnopravna Europi, a znanstvenoj domaćoj i stranoj javnosti svih struka – od prirodnih i tehničkih do društvenih i humanističkih – dalo pomagalo, građu i korpus za

razvoj strukovne (ali i opće) terminologije, te bilo
osnova za druga raznolika buduća istraživanja.</p>

</projectDesc>

</encodingDesc>

Slijedi primjer cijelog zaglavlja Mažuranić-Užarevićevog rječnika:

<teiHeader xml:lang="hr">

<fileDesc>

<titleStmt>

<title xml:lang="hr">Njemačko-ilirski slovar:
digitalno izdanje</title>

<title xml:lang="de">Deutsch-ilirisches
Wörterbuch: digitale Ausgabe</title>

<title xml:lang="en">German-Croatian dictionary:
digital edition</title>

<author>

<persName>

<forename>Ivan</forename>

<surname>Mažuranić</surname>

</persName>

</author>


```
<author>

<persName>

<forename>Jakov</forename>

<surname>Užarević</surname>

</persName>

</author>

<editor>

<persName>

<roleName>prof. dr. sc.</roleName>

<forename>Damir</forename>

<surname>Boras</surname>

</persName>

</editor>

<editor>

<persName>

<roleName>doc. dr. sc.</roleName>

<forename>Nikola</forename>

<surname>Ljubešić</surname>

</persName>
```

```

</editor>

<editor>

<persName>

<forename>Petra</forename>

<surname>Bago</surname>

</persName>

</editor>

<principal>

<persName>

<roleName>prof. dr. sc.</roleName>

<forename>Damir</forename>

<surname>Boras</surname>

</persName>

</principal>

</titleStmt>

<publicationStmt>

    <publisher>Filozofski fakultet Sveučilišta u
Zagrebu</publisher>

<address>

```

<addrLine>Filozofski fakultet Sveučilišta u Zagrebu</addrLine>

<addrLine>Ivana Lučića 3</addrLine>

<addrLine>10000 Zagreb</addrLine>

<addrLine>Hrvatska</addrLine>

</address>

<address xml:lang="en">

<addrLine>Faculty of Humanities and Social Sciences</addrLine>

<addrLine>Ivana Lučića 3</addrLine>

<addrLine>10000 Zagreb</addrLine>

<addrLine>Croatia</addrLine>

</address>

<pubPlace>Zagreb</pubPlace>

<date when="2014">2014</date>

</publicationStmt>

<notesStmt>

<note>Ovaj dokument pripremljen je u okviru znanstvenog projekta "Hrvatska rječnička baština i hrvatski europski identitet" (130-1301679-1380)

glavnog istraživača <persName><roleName>prof. dr. sc.</roleName><forename>Damira</forename><surname>Borasa</surname></persName>. Financiran je od strane Ministarstva znanosti, obrazovanja i sporta Republike Hrvatske <date from="2007" to="2013">od 2007. do 2013. godine.</date></note>

<note xml:lang="en">This document is prepared as part of the "Croatian dictionary heritage and Croatian European identity" scientific project (130-1301679-1380) of the main researcher <persName><roleName>prof.</roleName><forename>Damir</forename><surname>Boras</surname>,<roleName>PhD</roleName></persName>. This project is funded by the Croatian Ministry of Science and Technology <date from="2007" to="2013">from 2007. until 2013.</date></note>

</notesStmt>

<sourceDesc>

<bibl>

<title type="main" xml:lang="de"><choice><orig>Deutfch</orig><reg>Deutsc h</reg></choice>-<choice><orig>ilirifches</orig><reg>ilirisches</reg></choice> Wörterbuch</title>

<title type="main"
xml:lang="hr"><choice><orig>NĚMAČKO</orig><reg>NJEMAČ
KO</reg></choice>-ILIRSKI SLOVAR</title>

<author>

<persName>

<forename>Ivan</forename>

<surname>Mažuranić</surname>

</persName>

</author>

<author>

<persName>

<forename>Jakov</forename>

<surname>Užarević</surname>

</persName>

</author>

<publisher xml:lang="de">Verlag und Druck der
<choice><abbr>k.</abbr><expansion>königlichen</expansion></ch
oice><choice><abbr>priv.</abbr><expansion>privilegierten<
/expansion></choice><choice><orig>ilirifchen</orig><reg>i
lirischen</reg></choice> National-Buchdruckerei von
<persName><roleName>Dr.</roleName><forename>Ljudewit<

/forename><surname>Gaj</surname></persName></publishe
r>

<publisher xml:lang="hr"> Naklada i tisak
kraljevske privilegirane ilirske nacionalne
knjigotiskare

<persName><roleName>Dra.</roleName><forename>Ljudevit
a</forename><surname>Gaja</surname></persName></publi
sher>

<pubPlace xml:lang="de">Agram</pubPlace>

<pubPlace xml:lang="hr">Zagreb</pubPlace>

<date when="1842">1842</date>

</bibl>

</sourceDesc>

</fileDesc>

<encodingDesc>

<projectDesc>

<p>Od <date when="1595">1595. g.</date> kad je
objavljen prvi tiskani hrvatski rječnik
<persName><forename>Fausta</forename><surname>Vrančić
a</surname></persName> pa <date notBefore="1900"
notAfter="1925">do početka 20. st.</date> u Hrvatskoj
je objavljeno preko 300 različitih jednojezičnih,
dvojezičnih, višejezičnih, enciklopedijskih,

tehničkih, vojnih i drugih rječnika koji predstavljaju neprocjenjivu hrvatsku rječničku baštinu od kojih je većina nedostupna za javnu upotrebu. Upravo su stari hrvatski rječnici središte ovog znanstvenog projekta pod nazivom "Hrvatska rječnička baština i hrvatski europski identitet" (130-1301679-1380). Samo je ograničen broj tih rječnika strojno obrađen u znanstvenom projektu 130464 "Hrvatska rječnička baština i prikaz rječničkoga znanja" (<date from="2002" to="2005">2002-2005</date>) na koji se ovaj projekt nastavlja i čiji su rezultati djelomično objavljeni na "Portal hrvatske rječničke baštine". Oba projekta financiralo je Ministarstvo znanost, obrazovanja i sporta Republike Hrvatske, a glavni istraživač bio je <persName><roleName>prof. dr.</roleName><forename>Damir</forename><surname>Bora s</surname></persName>.</p>

<p>Nastavljajući se na dotadašnji rad, cilj projekta bio je dodatno istražiti hrvatsku rječničku baštinu proširivši istraživanje <date notAfter="1945">do 1945. g.</date> Taj iznenađujuće velik broj rječnika objavljen u velikom broju europskih gradova (u prvim stoljećima uglavnom izvan hrvatskih prostora) potvrđuje pretpostavku da su Hrvatska i hrvatski jezik (u svim tadašnjim nazivima i idiomima) bili prisutni u Europi, da su bili

neodvojivi dio europskog identiteta te da su svojim sadržajem povratno utjecali na svijest o hrvatskom identitetu u Europi.</p>

<p>Cilj je projekta bio i dodatno ispitati recepciju tih rječnika u nas i u Europi, utvrditi razloge njihova objavljivanja, dopuniti izrađenu bibliografiju te dalje popuniti postojeću digitalnu bazu slika i tekstova najvažnijih hrvatskih starih dvojezičnih i višejezičnih rječnika te razviti računalne metode za dodatnu rječničku, jezičnu, gramatičku, terminološku analizu rječničkih tekstova, kao i usporednu i temporalnu analizu između pojedinih rječnika. Cilj je bio i istražiti strukturu rječničkoga znanja sadržanoga u njima s posebnim naglaskom na stručnu terminologiju te razviti strojne sustave koji bi omogućili takvu analizu za hrvatski jezik. Na znanstveno utemeljenim načelima se prezentira tekstualna i slikovna baza izvornih hrvatskih rječničkih tekstova i njihov europski identitet.</p>

<p>Projekt je za Hrvatsku važan u znanstvenom, kulturnom i političkom kontekstu jer je pokazao da je Hrvatska u području leksikografije i grafičke rječničke prezentacije bila ravnopravna Europi, a znanstvenoj domaćoj i stranoj javnosti svih struka – od prirodnih i tehničkih do društvenih i humanističkih – dalo pomagalo, građu i korpus za

razvoj strukovne (ali i opće) terminologije, te bilo osnova za druga raznolika buduća istraživanja.</p>

</projectDesc>

</encodingDesc>

</teiHeader>

6.2.1.2 *Natuknice*

Na primjerima natuknica iz della Bellinog rječnika prikazat ćemo shemu korištenu za označavanje rječničkih natuknica.

Kao što je prethodno navedeno, za označavanje natuknice odabran je element <entry>. Element <form> označava sve oblike rječničke natuknice, pisane i govorene. Atributom *type* može se odrediti o kojem je obliku riječ. Prilikom označavanja rječnika same rječničke natuknice (*headword*) korišten je atribut *type*="lemma", koji označava da je natuknica u osnovnom obliku, što je u većini slučajeva i točno. Dodatnim se atributom definira jezik natuknice, važan podatak za dvojezične i višejezične rječnike. Budući da odabrani rječnici imaju samo pisanu varijantu natuknice, a nedostaje im izgovorna varijanta, korišten je element <orth>.

Element koji se može naći u svim dijelovima TEI dokumenta jest <pc>, koji označava interpunkcijske znakove. U zaglavlju nije upotrijebljen, ali je kod označavanja natuknica izričito označen.

Prema gornjem opisu, natuknica *Abate* izgledala ovako:

<form type="lemma" xml:lang="it">

```
<orth>Abate</orth>
```

```
<pc>.</pc>
```

```
</form>
```

Unutar natuknice može se nalaziti i oblik natuknice koji nije u osnovnom obliku, već u fleksiji. Uglavnom se u rječniku ne nalazi cijeli oblik riječi u fleksiji (`type="inflected"`), već samo nastavci. Razlog tomu je ušteda prostora. Tako je genitiv jednine imenice *Opat* prikazan samo kao *ta*, a označili bi ga na sljedeći način:

```
<form type="inflected">
```

```
<gramGrp>
```

```
<case value="genitive"/>
```

```
<number value="singular"/>
```

```
ta<pc>.</pc>
```

```
</gramGrp>
```

```
</form>
```

Ovdje se pojavljuju tri nova elementa. Element `<gramGrp>` grupira morfosintaktičke informacije jednog pojma. Element `<case>` navodi u kojem je padežu imenica te njegovu vrijednost putem atributa (`value="genitive"`), dok element `<number>` navodi broj imenice te njegovu vrijednost putem atributa (`value="singular"`). Budući da posljednja dva podatka nisu izričito izrečeni, već samo implicitno putem nastavka *ta*, dodani su kao prazni elementi bez sadržaja. Kao što je već

prethodno spomenuto, sadržaji elemenata biti će samo oni dijelovi koji se nalaze u izvorniku.

Nadalje, elementom `<form>` može se označiti i složena natuknica (`type="compound"`) koja se sastoji od jednostavnih leksičkih jedinica, od kojih je jedna rječnička natuknica (*headword*) u osnovnom obliku. Element `<oRef/>` služi za referiranje na natuknicu u osnovnom obliku. Na primjer, u natuknici Abate nalazi se složena natuknica Dignita d'Abate:

```
<form type="compound">
  <orth>Dignità d'<oRef/></orth>
  <pc>.</pc>
</form>
```

U rječniku se nakon nastavka genitiva jednine kod imenice nalazi oznaka za rod (`<gen>`), a vrijednost roda može se zabilježiti atributom (`value="m"`). Osim oznake za rod, dodana je i oznaka za vrstu riječi (`<pos>`), a vrijednost vrste riječi može se zabilježiti atributom (`norm="noun"`). Oznaka za vrstu riječi je prazan element jer nije izričito izrečeno u rječniku. Na primjer, nakon već spomenute imenice Opat nalazi se nastavak za genitiv jednine *ta* te oznaka za muški rod:

```
<gramGrp>
  <pos norm="noun"/>
  <gen value="m">m<pc>.</pc></gen>
</gramGrp>
```

Element `<sense>` grupira različita značenja natuknice, kao što su definicije, primjeri i prijevodi. Atributom `n` može se označiti broj različitih značenja. Unutar tog elementa, elementom `<cit>` označuju se prijevodi (`type="translation"`) natuknice (*headword*) te primjeri (`type="example"`). Obavezan element unutar elementa `<cit>` je `<quote>`. Tako je prijevod natuknice *Abate* na latinskom ovako označen:

```
<cit type="translation" xml:lang="la">
  <quote>Abbas<pc>,</pc></quote>
</cit>
```

Primjeri hrvatskih prijevoda u *della Bellinu* rječniku označeni su na sljedeći način:

```
<cit type="example" xml:lang="hr">
```

dok se unutar prijevoda nalazi i oznaka izvora primjera označena elementom `<bibl>`. Na primjer:

```
<cit type="example" xml:lang="hr">
  <quote>Evo gre pet godin', dàsam gne sluga
ja<pc>;</pc></quote>
  <bibl>Scifc<pc>.</pc></bibl>
</cit>
```

Ponekad u rječnicima postoji uputa na koje se značenje riječi misli. Takve upute označene su elementom `<usg>` i atributom `type="hint"`:

```
<usg type="hint">
```

```
per
```

```
metter
```

```
radici
```

```
<pc>.</pc>
```

```
</usg>
```

U rječnicima su česte uputnice na druge natuknice. One su označene elementom `<xr>` i atributom `type="see"`. Unutar tog elementa nalazi se element `<ref>` kojim je označena natuknica na koju se referira. Budući da taj element upućuje na natuknice unutar rječnika, ali i na vanjske resurse, potrebno je definirati lokaciju na koju se upućuje. To je moguće atributom `target`, gdje se u vrijednost atributa stavi lokacija reference, kao što je to u ovom primjeru:

```
<xr type="see">V.
```

```
<ref target="#barbare">Barba-re</ref>
```

```
<pc>.</pc>
```

```
<ref target="#radicare">Radicare</ref>
```

```
<pc>.</pc>
```

```
</xr>
```

Konačno, označeni su i prelasci u novi red (`<lb/>`), u novi stupac (`<cb/>`) i na novu stranicu (`<pb/>`).

Prema ovim TEI oznakama načinjen je pojednostavljen popis oznaka za strukturu (19), koje su objašnjene u poglavlju 6.3.2.2.

6.2.2 Primjer natuknice cijele označene TEI oznakama

Slijedi primjer cijele natuknice *Abate* iz *della Bellinog* rječnika označene TEI oznakama.

```
<entry>

<form type="lemma" xml:lang="it">

<orth>Abate</orth>

<pc>.</pc>

</form>

<sense n="1">

<cit type="translation" xml:lang="la">

<quote>Abbas<pc>,</pc></quote>

<form type="inflected">

<gramGrp>

<case value="genitive"/>

<number value="singular"/>

tis<pc>.</pc>

</gramGrp>
```

```

</form>

<gramGrp>

<pos norm="noun"/>

<gen value="m">m<pc>.</pc></gen>

</gramGrp>

</cit>

<cit type="translation" xml:lang="hr">

<quote>Opat<pc>,</pc></quote>

<form type="inflected">

<gramGrp>

<case value="genitive"/>

<number value="singular"/>

    ta<pc>.</pc>

</gramGrp>

</form>

<gramGrp>

<pos norm="noun"/>

<gen value="m">m<pc>.</pc></gen>

</gramGrp>

```

```

</cit>

<cit type="translation" xml:lang="hr">

<quote>Igu-<lb/>men<pc>,</pc></quote>

<form type="inflected">

<gramGrp>

<case value="genitive"/>

<number value="singular"/>

    ena<pc>.</pc>

</gramGrp>

</form>

<gramGrp>

<pos norm="noun"/>

<gen value="m">m<pc>.</pc></gen>

</gramGrp>

</cit>

</sense>

<sense n="2">

<form type="compound">

<orth>Dignità d'<oRef/></orth>

```



```

<pc>.</pc>

</form>

<cit type="translation" xml:lang="hr">

<quote>Opat-<lb/>ftvo<pc>,</pc></quote><!--
primjer komentara -->

<form type="inflected">

<gramGrp>

<case value="genitive"/>

<number value="singular"/>

    va<pc>.</pc>

</gramGrp>

</form>

<gramGrp>

<pos norm="noun"/>

<gen value="n">n<pc>.</pc></gen>

</gramGrp>

</cit>

<cit type="translation" xml:lang="hr">

<quote>Igumenftvo<pc>,</pc></quote>

```

```
<form type="inflected">

<gramGrp>

<case value="genitive"/>

<number value="singular"/>

    tva<pc>.</pc>

</gramGrp>

</form>

<gramGrp>

<pos norm="noun"/>

<gen value="m">m<pc>.</pc></gen>

</gramGrp>

</cit>

</sense>

</entry>
```

6.3. Primjena poluautomatskih metoda za opis shemom

6.3.1. Osnovni pojmovi strojnog učenja

Područje strojnog učenja (*machine learning*) interdisciplinarno je područje koje preuzima metode i tehnike iz računarstva, umjetne inteligencije, psihologije, teorije informacija, statistike i vjerojatnosti, filozofije, neuroznanosti, genetike i dr. Strojno učenje možemo definirati kao računalne metode koje koriste iskustvo kako bi poboljšale performanse ili napravile točna predviđanja, gdje se iskustvo uglavnom odnosi na digitalne podatke prikupljene i dostupne za analizu. Cilj strojnog učenja je razviti učinkovite i točne algoritme za predviđanje.¹⁶⁹

Prema Mitchellu¹⁷⁰, računalni program uči iz iskustva I s obzirom na neku vrstu zadatka Z i mjeru učinkovitosti U , ako se njegova učinkovitost pri zadatku Z , mjerena s U , poboljšava s iskustvom I . Pojednostavljeno, cilj je stvoriti sustav čiji se rezultati poboljšavaju kroz iskustvo.

U istraživanju je primijenjena metoda strojnog učenja za označavanje jezika i strukture rječničkih natuknica. Prema prethodno navedenoj definiciji možemo odrediti elemente strojnog učenja (zadatak Z , mjera učinkovitosti U , iskustvo I). U slučaju označavanja jezika vrijedi sljedeće:

- zadatak Z : odrediti jezik određene pojavnice
- mjera učinkovitosti U : postotak točno dodijeljenih jezičnih oznaka

¹⁶⁹Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet. Foundations of Machine Learning. MIT Press, 2012, (str. 1).

¹⁷⁰Mitchell, Tom M. Machine learning. New York : McGraw Hill, 1997, (str. 2).

- iskustvo I : ručno označeni skup za učenje.

U slučaju označavanja strukture vrijedi sljedeće:

- zadatak Z : odrediti kojem dijelu strukture pripada pojava prema definiranoj shemi
- mjera učinkovitosti U : postotak točno dodijeljenih strukturnih oznaka
- iskustvo I : ručno označeni skup za učenje.

Strojno učenje može se primijeniti pri rješavanju raznih problema. Osnovne vrste zadataka kod kojih se može primijeniti strojno učenje su klasifikacija (*classification*), regresija (*regression*), rangiranje (*ranking*), grupiranje (*clustering*) i smanjenje dimenzionalnosti (*dimensionality reduction*). Kod klasifikacije je cilj dodijeliti kategoriju svakom entitetu, dok se kod regresije svakom entitetu dodjeljuje realna vrijednost. Rangiranje se bavi redanjem entiteta prema određenom kriteriju. Cilj grupiranja je dijeljenje entiteta u homogene grupe. Konačno, smanjenje dimenzionalnosti pretvara početni prikaz entiteta u prikaz s manje dimenzija zadržavajući neka svojstva početnog prikaza.¹⁷¹

Osnovni pristupi strojnom učenju su nadzirano (*supervised learning*), nenadzirano (*unsupervised learning*) i polunadzirano učenje (*semi-supervised learning*). Kod nadziranog učenja algoritmu je dan skup označenih podataka, odnosno skup za učenje (*training set*), na kojem algoritam uči te vrši predviđanja na prethodno neviđenim podacima. Klasifikacija, regresija i rangiranje pripadaju nadziranom pristupu učenja. Kod nenadziranog učenja algoritmu je dan skup neoznačenih podataka te vrši predviđanja na prethodno

¹⁷¹Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet. Foundations of Machine Learning. MIT Press, 2012, (str. 2).

neviđenim podacima. Grupiranje i smanjenje dimenzionalnosti pripadaju nenadziranom učenju. Kod polunadziranog učenja skup za učenje sastoji se od označenih i neoznačenih podataka, na temelju kojih algoritam vrši predviđanja na prethodno neviđenim podacima. Ovaj je pristup čest u slučajevima kada su neoznačeni podaci lako dostupni, ali je označavanje otežano.¹⁷²

Osnovne termine povezane s nadziranim strojnim učenjem objasniti ćemo u nastavku. **Primjeri** ili **instance** (*instances*) su podaci koje koristimo za učenje i evaluaciju¹⁷³, što u ovom istraživanju predstavlja pojavnice nasumično odabranih natuknica koje čine ručno označeni skup za učenje. Mitchell to naziva iskustvo *I*. **Značajke** (*features*) su skup atributa pridruženih primjeru ili instanci koji predstavljaju određenu formalizaciju tog primjera.¹⁷⁴ Tako pojavnici u zadatku označavanja jezika predstavljamo značajkama samog oblika pojavnice, oblika prethodnih i sljedećih pojava, trigramima znakova i sl. **Atributi** su vrijednosti tih značajki. U ovom istraživanju pokušavamo identificirati one značajke koje daju najbolje rezultate nad našim zadacima (npr. postižu li se bolji rezultati s pojavnicom zapisanom u izvornom obliku ili pretvorenom u kurent (mala slova)). Taj se postupak naziva **odabirom značajki** (*feature selection*). **Oznake** (*label*) su vrijednosti ili kategorije dodijeljene primjerima ili instancama¹⁷⁵, odnosno to su vrijednosti koje pokušavamo predvidjeti nad neviđenim podacima. U statistici se značajke instanci često nazivaju **nezavisne varijable** (*explanatory variables*), dok se oznake nazivaju **zavisnom**

¹⁷²Ibid. (str. 7).

¹⁷³Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet. Foundations of Machine Learning. MIT Press, 2012. (str. 3).

¹⁷⁴Ibid. (str. 3).

¹⁷⁵Ibid. (str. 3).

varijablom (*response variable*). Osnovni cilj nadziranog strojnog učenja jest izmodelirati zavisnost zavisne varijable od nezavisnih te omogućiti predviđanje vrijednosti zavisne varijable na temelju vrijednosti nezavisnih na do tada neviđenim podacima. Budući da koristimo metodu klasifikacije, u ovom istraživanju dodijeljene su tri kategorije za jezik i 19 kategorija za strukturu. **Skup za učenje** (*training sample*) služi za učenje algoritma na temelju instanci s pripadajućim oznakama.¹⁷⁶ **Skup za validaciju** (*validation sample*) čine instance kojima se usklađuju parametri algoritma za učenje za vrijeme rada na označenim podacima.¹⁷⁷ Količina označenih podataka često bude premala da bi se dio mogao izdvojiti za skup za validaciju, jer tada ne bi bilo dovoljno podataka u skupu za učenje, što je i slučaj u ovom istraživanju. Upravo se zato često, što je slučaj i u ovom istraživanju, primjenjuje metoda ***n*-struke unakrsne validacije**, koja je objašnjena u poglavlju 6.3.3.3. Uzorak za učenje čine instance kojima se naizmjenično gradi model za predviđanje te vrši evaluacija tog modela.¹⁷⁸

Bitna karakteristika teksta jest da se pojavnice nalaze u određenom međuodnosu, odnosno da nisu nasumično poredane. Ako se metode strojnog učenja upotrebljavaju nad tekstom, gdje su pojavnice najčešće instance, činjenica jest da su te instance u međuodnosu. Kada se instance nalaze u određenom međuodnosu, znači da su podaci kojima raspolažemo **sekvencijalni podaci**, odnosno **sekvence**. Dvije osnovne tehnike učenja iz sekvencijalnih podataka su skriveni Markovljevi modeli (*HiddenMarkov models*, *HMM*) i

¹⁷⁶Ibid. (str. 3).

¹⁷⁷Ibid. (str. 3).

¹⁷⁸Ibid. (str. 4).

uvjetna nasumična polja (*Conditional random fields, CRF*).¹⁷⁹ HMM je statistički model u kojem se za modelirani sustav pretpostavlja da je Markovljev proces sa skrivenim stanjima.¹⁸⁰ Markovljev proces možemo definirati kao stohastički proces u kojem uvjetna distribucija vjerojatnosti budućih stanja procesa, uz zadana sadašnja i sva prošla stanja, ovisi samo o sadašnjem stanju.¹⁸¹ CRF je statistički model koji se često koristi za predviđanje strukture, te kao i HMM pripada probabilističkim grafičkim modelima. Općenito se CRF može shvatiti kao generalizaciju HMM-a.¹⁸² Skriveni Markovljevi modeli pripadaju generativnim algoritmima, dok su njihov pandan kod diskriminativnih algoritama uvjetna nasumična polja. Osnovna razlika između te dvije vrste algoritama je da diskriminativni algoritmi ne mogu generirati uzorke iz zajedničke distribucije ulaznih i izlaznih varijabli, dok generativni mogu. Budući da kod klasifikacije nije potrebna zajednička distribucija, diskriminativni modeli mogu postići bolje rezultate.¹⁸³

Od sedam rječnika nad kojima je provedena prva faza istraživanja, opisana druga faza provedena je samo nad prvim sveskom della Bellinog

¹⁷⁹**Lafferty, J., McCallum, A., Pereira, F.** (2001). "Conditional random fields: Probabilistic models for segmenting and labeling sequence data". Proc. 18th International Conf. on Machine Learning. Morgan Kaufmann. pp. 282-289.

¹⁸⁰**Hidden Markov Models.** // Encyclopedia of Machine Learning. Sammut, Claude; Webb, Geoffrey I. (ur.). New York : Springer, 2011.

¹⁸¹**Markov Process.** // Encyclopedia of Machine Learning. Sammut, Claude; Webb, Geoffrey I. (ur.). New York : Springer, 2011.

¹⁸²**Conditional random field.** // Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Conditional_random_field (26. 9. 2014.).

¹⁸³**Discriminative model.** // Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Discriminative_model (26. 9. 2014.).

6.3.2. Skup podataka

Prvi svezak sadrži 7 972 rječničkih natuknica od slova A do H (Huquang), koje su automatski opojavničene. U pojavnice općenito ulaze slova, brojke i ostali pravopisni znakovi.¹⁸⁴ Svezak sveukupno sadrži 403 128 pojavnica. U nastavku je detaljnije objašnjeno što je sve određeno kao samostalna pojavnica.

6.3.2.1 Pojavnice

¹⁸⁴**Boras, Damir.** "Teorija i pravila segmentacije teksta na hrvatskom jeziku." Doktorska disertacija. Filozofski fakultet, Sveučilište u Zagrebu, 1998.

¹⁸⁵U della Bellinom rječniku nalaze se sljedeće interpunkcije: „ , , , , , , , !“ , .“ .“ “ (, , ,)“ .

U della Bellinom rječniku se ispred i iza interpunkcijskih znakova uvijek nalazi razmak, osim u slučaju interpunkcijskog znaka točke. U slučaju točke, ako se ispred nje nalazi razmak, riječ je o pravoj interpunkciji, a ako nema razmaka, radi se o kraticama. U rječniku je naveden popis mogućih kratica (vidi u Prilogu). Međutim, tijekom ručnog označavanja uočena su odstupanja od popisa. U većini se slučajeva radi o inačicama postojećih kratica:

- fem. umjesto F., Femm.
- Gjorg. Pf. umjesto Gjorg. Ps.
- Gjor. Pfal. umjesto Gjorg. Salm.
- Kafs. umjesto Kass.
- Ofm. umjesto Osm.
- Pal. Chr. umjesto Palm. Chr.
- plur. umjesto pl.
- Scif. umjesto Scis.
- fuft. umjesto Suft.
- V. Bafs. umjesto V. Baff.

Nadalje, uočene je jedna kratica koja se ne nalazi u popisu, a to je:

- Defcr.

Ipak, kad je riječ o točki kao pravoj interpunkciji, ima primjera pogrešaka u rječniku gdje točka ne završava dozvoljenu kraticu ili identificiranu kraticu.

Sve navedeno dobro je ilustrirano sljedećima primjerima.

Na primjeru natuknica *Abbaglio* i *Abbeveratore* vidljiva je razlika između točke kao općenitog (pravog) interpunkcijskog znaka i točke na kraju kratice. U natuknici *Abbaglio*:

Abbaglio , errore , V. Abbagliamento .

nalaze se dvije točke: Prva točka pripada kratici za uputnicu (V. – *vedi* odnosno vidi) i nema razmaka između slova i točke. U ovom slučaju pojavnicu čini kratica s točkom. Druga točka pripada kraju natuknice i ima razmak između riječi i točke te čini zasebnu pojavnicu.

U slučaju natuknice Abbeveratore:

Abbeveratore . <i>Potionator , ris . m.</i>

Opoitegl , ◇glja . n.

možemo naći pet točaka. Prva označava kraj talijanskog dijela natuknice, nakon čega počinje latinski prijevod. Druga označava kraj latinskog dijela natuknice, dok je treća točka sastavni dio kratice za oznaku imenice muškog roda pa nema razmaka između teksta i točke. Četvrta točka je oznaka za kraj hrvatskog prijevoda. Peta točka je sastavni dio kratice za oznaku imenice srednjeg roda. U ovom se slučaju zasebnim pojavnicama smatraju prva, druga i četvrta točka, dok su treća i peta točka sastavni dio kratice, te ne čine zasebne pojavnice.

Bitno je spomenuti da u rječniku možemo naći vrlo malen broj neispravnosti kao što je slučaj u natuknici Abbiettiffimo:

Abbiettiffimo . <i>Abiectiffimus , a .

m.</i> Najpoti-|fctenji , ija , ije .

Pripotifcten , na , no .◇Najpripsoftji

, ija , ije .

Neispravnost se nalazi u latinskom dijelu natuknice, gdje oznaka m. upućuje na muški rod (nepostojeće imenice) umjesto da ispravno bude razmak ispred točke koji tada označava nastavak srednjeg roda pridjeva. Ispravan latinski dio natuknice izgledao bi ovako:

Abjectiffimus , a . m .

Takve neispravnosti uobičajena su pojava u starim rječnicima zato što je slog rječnika ručno slagan, pa je točnost ovisila i o koncentraciji, odnosno umoru ili „raspoloženju“ („stanju“) slagara.

Takve neispravnosti prema predviđenom modelu bit će kao takve zadržane kao i u originalu u digitaliziranom tekstu rječnika, ali i prikladno označene.

Neispravan oblik bit će „izrijekom“ označen kao pogreška, kako bi se ostavila mogućnost analize pogrešaka te povisiti točnost drugih analiza (npr. bit će jasna razlika između latinskog nastavka za srednji rod pridjeva i kratice za muški rod imenice). Ujedno će se modelom omogućiti i prikaz rječničke inačice sa svim uočenim, ispravljenim pogreškama.

Međutim, ručno označavanje i ispravljanje pogrešaka izvan je područja interesa ovog rada, ali je jasno da automatsko označavanje omogućuje i jednostavnije lociranje mogućih pogrešaka.

Nadalje, zbog sažimanja prostora, u rječniku se često nalaze samo nastavci za stvaranje drugih oblika riječi, a ne cijele riječi. Zbog toga u pojavnice ulaze i imenični nastavci za genitiv jednine, a ponegdje možemo naći i nastavak za genitiv množine.

To se dobro vidi u natuknici Amello:

Amello , erba . *Amellus* .
Divglifciza , ize .f. Rutta divja ,
vje . f.

gdje možemo vidjeti dva imenična nastavka za genitiv jednine: ize i vje.

U natuknici pak Anelletti:

Anelletti , a' quali fi affibbiano gli
uncinelli delle vesti . <i>Anuli
vestiarii .</i> Spone , naa .f. plur.

možemo vidjeti nastavak za imenični genitiv množine, što je i vidljivo iz kratice plur. na kraju natuknice.

Od pridjevskih oblika nalazimo nastavke za ženski i srednji rod, dok se muški oblik nalazi na početku prijevoda u potpunom obliku. Ti oblici se također nalaze kao zasebne pojavnice u digitalnom korpusu rječnika.

Primjer za latinski i hrvatski dio možemo naći u već spomenutoj natuknici Abbiettiffimo:

Abbiettiffimo . <i>Abiectiffimus , a .
m.</i> Najpoti-|fctenji , ija , ije .
Pripotifcten , na , no .f. plur. Najpriproftji
, ija , ije .

Od glagolskih oblika možemo u hrvatskom prijevodu naći nastavke za prvo lice jednine sadašnjeg te prvo lice jednine prošlog vremena, koji čine zasebne pojavnice u korpusu. Primjere možemo vidjeti u natuknici Anfiare:

Anfiare , ed anfare , respirare con
difficoltà .f. plur. Anhelo , as .</i>
Şapjehatife , pjuhàmfe , hòsàm-|fe .
Trudnòdihati , diham , hòfam . Pjeha-
|ti , ham , hòfam . Stegnati , ftegnem
, fte-|gnaòfam .

gdje su nastavci pjuhàmfe, diham i ham prvog lica jednine sadašnjeg vremena, a nastavci hòsàm-|fe, hòfam i hòfam prvog lica jednine prošlog vremena. U zadnjem prijevodu su ftegnem i fte-|gnaòfam potpuni oblici prvog lica jednine sadašnjeg i prošlog vremena glagola Stegnati.

Kod latinskih glagola nalazimo uglavnom glagol u prvom licu jednine sadašnjeg vremena, nakon kojeg slijedi nastavak za infinitiv. Primjer možemo vidjeti u natuknici *Anfiare*:

Anfiare , ed anfare , reſpirare con
difficoltà .<i>Anhelo , as .</i>
ſapjehatife , pjeħàmfe , hòsàm-|fe .
Trudnòdihati , diham , hòfam . Pjeħa-
|ti , ham , hòfam . Stegnati , ftegnem
, fte-|gnaòfam .

gdje je nastavak za infinitiv „as“. U rjeđim slučajevima je nakon nastavka za infinitiv još naveden nastavak za prvo lice jednine prošlog vremena te glagolski pridjev u srednjem rodu, kao što je u natuknici *Adagiare*:

Adagiare , agiare , accomodare . <i>Curo ,
as ,<avi , atum .</i> Naftaniti dobro ,
ftanujem ,<ftaniòfam . Pofaditi dobro .
V. Accomodare .

Nastavci su sljedeći: nastavak za infinitiv je „as“, nastavak za prvo lice jednine prošlog vremena je „avi“ i nastavak za glagolski pridjev u srednjem rodu je „atum“. Svi navedeni nastavci čine zasebne pojavnice u korpusu.

U digitaliziranom korpusu postoje posebni znakovi koji omogućuju dodatno prepoznavanje dijelova natuknica. Budući da su uneseni već samom digitalizacijom, a korisni su za prepoznavanje strukture natuknice, zadržani su i u bazi digitalnih podataka te označeni posebnim znakovima. I oni u formalnom smislu predstavljaju pojavnice. U tu skupinu ulaze dakle još i oznake za prelazak u novi red unutar iste natuknice, prelazak u novi stupac te prelazak na novu stranicu. Te oznake predstavljaju zasebne pojavnice u slučaju da ne

razdvajaju riječ na dva dijela. Kao oznaka za prelazak u novi red unutar iste natuknice izabran je znak \diamond što se dobro vidi u natuknici Abbeveratore:

Abbeveratore . *<i>Potionator , ris . m.</i>*
Opoitegl , \diamond glja . n.

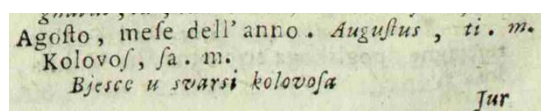
Oznaka za prelazak u novi stupac je „*<cb>*“ i vidljiva je u natuknici Abbreviatrice:

Abbreviatrice . Skratitegliza , ze . f.
Pokra-|titegliza , ze . f. Skracjaliza
, ze . f.*<cb>*

dok je oznaka za prelazak na novu stranicu „*<pb>*“ i nalazi se u natuknici Agofto:

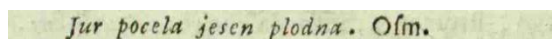
Agofto , mefe dell' anno . *<i>Auguftus , ti . m.</i>*
<i>. m.</i> \diamond Kolovoš , ša . m. \diamond *<x>* Bjesce u
svarsı kolovoša*<pb>* Jur pocela jesen
plodna .*</x>* Ofm.

Slika 6. 1 Primjer natuknice



Agofto , mefe dell' anno . *Auguftus , ti . m.*
Kolovoš , ša . m.
Bjesce u svarsı kolovoša

Jur



Jur pocela jesen plodna . Ofm.

Međutim, ako se prethodno navedene oznake nalaze unutar riječi, odlučili smo ih zadržati unutar riječi tako da s preostalim znakovima pojavnice čine samo jednu pojavnicu. Zato se, na primjer, u rječniku može naći 11 500

pojavnica u kojima se nalazi oznaka za prelazak u novi red, 87 pojavnica unutar kojih se nalazi prelazak u novi stupac i 83 pojavnice s oznakom za prelazak na novu stranicu.

Kao oznaka za prelazak u novi red unutar riječi uzet je znak | različit od oznake za prijelaz u novi red koji čini posebnu pojavnicu („◊“). Time se olakšava formalna obrada, ali se ujedno i čitateljima digitalnog teksta olakšava čitanje.

Kako to izgleda može se vidjeti u natuknici Abbiettiffimo:

Abbiettiffimo . <i>Abjectiffimus , a .
m.</i> Najpoti-|fctenji , ija , ije .
Pripotifcten , na , no .◊Najpriproftji
, ija , ije .

Primjer pojavnice unutar koje se nalazi oznaka za prelazak u novi stupac nalazi se u natuknici Accetofiffimo:

Accetofiffimo . <i>Accidiffimus , ma , mum
.</i> Pri-<cb>gljut , gljuta , gljuto .
Prioften , na , no .

dok se primjer pojavnice unutar koje se nalazi oznaka za prelazak na novu stranicu može naći u natuknici Adefcato:

Adefcato . <i>Inefcatus , a , um .</i>
Mamgljen , na ,◊no . V. Allettato .
Adefcato , già alletta-<pb>to , o prefo
all' efca . Domamgljen , na ,◊no .
Primamgljen , na , no . Dovabgljen ,◊na
, no .

Nadalje, od postojećih oznaka u pojavaice ulaze još i oznake za početak i kraj latinskih dijelova natuknica. Oznaka za početak latinskog dijela natuknice je „<i>“, a oznaka za kraj „</i>“, što je u standardnom HTML-u oznaka za kurziv (*italic*) jer su latinski tekstovi u rječniku uvijek i bez iznimke tiskani kurzivom. Primjer možemo naći u natuknici Abbeveratore:

```
Abbeveratore . <i>Potionator , ris . m.</i>  
Opoitegl , ðglja . n.
```

Međutim, ima i drugih tekstova tiskanih kurzivom, a to je i zanimljivost ovoga rječnika da se u njemu nalaze i citati iz djela koja su korištena kao korpus za izradu rječnika. Na kraju tako označenih citata kratice su izvornih dijela, navedenih u predgovoru rječnika, koje označavaju izvore citata. Zato u pojavaice ulaze još i oznake za početak i kraj citata, koje su već u korpusu pri digitalizaciji. Oznaka za početak citata je „<x>“, dok je oznaka za kraj „</x>“. Primjer možemo naći u natuknici Abbadeffa:

```
Abbadeffa , <i>Abbatiffa , æ . f.</i>  
Opattiza , ze . f.ð<x>Tuj gnu poftavj  
ša Majku , i ša Opattizu</x> Gjo.ðS.  
Ben. Gofpoghja od Manaftjera . Igume-  
lniza , ize . f.
```

Sami izvori citata nisu bili označeni, ali su kasnije u ručnom označavanju izričito istaknuti, jer se uglavnom nalaze iza oznake za kraj citata.

Ukratko, u pojavaice se ubrajaju:

1. riječi
2. riječi unutar kojih se nalazi oznaka za prelazak u novi red

3. riječi unutar kojih se nalazi oznaka za prelazak u novi stupac
4. riječi unutar kojih se nalazi oznaka za prelazak na novu stranicu
5. interpunkcijski znakovi
6. kratice
7. imenični nastavak za genitiv jednine
8. imenični nastavak za genitiv množine
9. pridjevski nastavak za ženski rod
- 10.pridjevski nastavak za srednji rod
- 11.nastavak za prvo lice jednine sadašnjeg vremena
- 12.nastavak za prvo lice jednine prošlog vremena
- 13.oznaka za prelazak u novi red unutar jedne natuknice („↑“)
- 14.oznaka za prelazak u novi stupac („→“)
- 15.oznaka za prelazak na novu stranicu („↓“)
- 16.oznaka za početak latinskog dijela natuknice („↔“)
- 17.oznaka za kraj latinskog dijela natuknice („↕“)
- 18.oznaka za početak citata („““)
- 19.oznaka za kraj citata („”“).

Prema navedenom popisu pravila za određivanje pojava, prvi svezak della Bellinog rječnika sadrži sveukupno 403 128 pojava. To znači da u prosjeku jedna rječnička natuknica sadrži 50,57 pojava.

6.3.2.2 Skup za učenje

Nakon što su natuknice opojavničene, iz korpusa je slučajnim odabirom izdvojena 101 natuknica za ručno označavanje. Popis izdvojenih natuknica nalazi se u prilogu. Ručno označavanje vršio je stručni označitelj koji dobro poznaje della Bellin rječnik, odnosno u ovom slučaju autorica disertacije.

Svaka pojava unutar 101 izdvojene natuknice označena je na dvije razine: oznakom za jezik i oznakom za strukturu. Jezična razina ima tri različite oznake, dok strukturna razina ima 19 oznaka. Distribucije obaju skupina oznaka prikazane su u Tablici 6.1 i Tablici 6.2. Sveukupno je ručno označeno 8 340 oznaka u svakoj skupini, odnosno sveukupno 16 680 oznaka. To znači da u prosjeku jedna rječnička natuknica u ručno označenom uzorku sadrži 82,57 pojava, odnosno 32 pojavnice više nego prosjek natuknice u cijelom prvom svesku rječnika.

Kratice korištene za jezične oznake te njihovo tumačenje su sljedeće:

- hr – označava pojavu na hrvatskom jeziku
- it – označava pojavu na talijanskom jeziku
- la – označava pojavu na latinskom jeziku.

Tablica 6.1 Distribucija jezičnih oznaka

| jezik | frekvencija |
|--------------|--------------------|
| hr | 4 395 |
| it | 2 164 |
| la | 1 781 |

Zanimljiv je podatak da je preko polovine (53 %) pojava označeno hrvatskim jezikom, dok je talijanski jezik nešto više zastupljen nego latinski (26 % naspram 21 %), što pokazuje temeljitost autora u prikazu svih mogućih riječi sličnih značenja u hrvatskom jeziku u ovom rječniku, a time i bogatstva hrvatskoga jezika. Jasno je da je zbog bliskosti talijanskoga i latinskoga jezika, većinom kao latinski ekvivalent talijanskoga pojma uzeta samo jedna latinska riječ, dok je za hrvatski u prosjeku (prema ručnom označavanju) navedeno gotovo dvije i pol riječi (2,47), što znači da su u prosjeku navedena podjednako ili dva ili tri ekvivalenta na hrvatskom jeziku.

Strukturna razina temelji se na TEI oznakama opisanim u poglavlju 6.2.1.2 te ima 19 oznaka. Korištene kratice te njihova tumačenja su sljedeći:

1. `abbr` – označava kraticu
2. `adj` – označava nastavak za pridjev
3. `adjf` – označava nastavak za ženski rod jednine pridjeva
4. `adjn` – označava nastavak za srednji rod jednine pridjeva
5. `bibl` – označava izvor citata

6. `cb` – označava pojavnicu za prelazak u novi stupac u slučaju kada ne razdvaja riječ¹⁸⁶
7. `citex` – označava citat
8. `citttrans` – označava prijevod rječničke natuknice ili neke druge riječi unutar natuknice
9. `formlem` – označava rječničku natuknicu¹⁸⁷
10. `genpl` – označava nastavak za genitiv množine imenice
11. `gensg` – označava nastavak za genitiv jednine imenice
12. `hint` – označava pojavnicu koja pobliže usmjerava značenje rječničke natuknice ili neke druge riječi unutar natuknice
13. `lb` – označava pojavnicu za prelazak u novi red kada se nalazi unutar jedne natuknice i ne razdvaja riječ¹⁸⁸
14. `pb` – označava pojavnicu za prelazak na novu stranicu kada se ne nalazi unutar jedne riječi¹⁸⁹
15. `pc` – označava interpunkcijske znakove koji nisu dio kratice
16. `pos` – označava vrstu riječi (muški, ženski i srednji rod imenice, množinu u slučaju da se imenica nalazi u tom obliku, pridjev, prilog)

¹⁸⁶U slučaju kada se nalazi unutar jedne riječi, ostavljena je oznaka unutar te riječi te je zabilježeno kao jedna pojavnica. U ručno označenom korpusu nalaze se takva dva slučaja.

¹⁸⁷Budući da je ručno označena 101 natuknica, a u korpusu ima 125 pojavnica s oznakom `formlem`, potrebno je objasniti tu razliku. Razlika u brojevima nastala je iz dva razloga. Jedan razlog je da rječnička natuknica može imati više od jednog oblika, kao što je tu primjeru natuknice *Affrica*, gdje je naveden i dodatni oblik *Africa*. Oba oblika označena su oznakom `formlem`. Drugi razlog je da se rječnička natuknica može sastojati od jedne ili više pojavnica, kao što je slučaj s natuknicom *Coftipáto nel detto fenfo*, gdje su sve pojavnice također označene oznakom `formlem`.

¹⁸⁸U slučaju kada se nalazi unutar jedne riječi, ostavljena je oznaka unutar te riječi te je zabilježeno kao jedna pojavnica. U ručno označenom korpusu nalazi se 271 takav slučaj.

¹⁸⁹U slučaju kada se nalazi unutar jedne riječi, ostavljena je oznaka unutar te riječi te je zabilježeno kao jedna pojavnica. U ručno označenom korpusu nalaze se dva takva slučaja.

17. *ref* – natuknica na koju uputnica upućuje
18. *v* – označava nastavak za glagolski oblik, uglavnom prvo lice jednine sadašnjeg te prvo lice jednine prošlog vremena
19. *xr* – oznaka za uputnicu.

Tablica 6.2 Distribucija strukturnih oznaka

| struktura | frekvencija |
|------------------|--------------------|
| abbr | 55 |
| adj | 2 |
| adjf | 109 |
| adjn | 118 |
| bibl | 90 |
| cb | 7 |
| citex | 729 |
| cittrans | 3 167 |
| formlem | 125 |
| genpl | 1 |
| gensg | 185 |
| hint | 415 |
| lb | 506 |
| pb | 6 |
| pc | 2 329 |
| pos | 198 |
| ref | 35 |
| v | 230 |
| xr | 33 |

Ručno označeni korpus, odnosno skup za učenje algoritma, sadrži 8 340 pojava (2,07 %), dok skup neoznačenih podataka sadrži ostalih 394 788 (97,93 %).

6.3.2.3 *Primjer natuknice označene na jezičnoj i strukturnoj razini*

Istraživanje je provedeno u dvije faze. U prvoj se fazi rješavao problem označavanja svake pojavnice odgovarajućom oznakom za jezik. Druga se faza bavila rješavanjem problema označavanja svake pojavnice odgovarajućom oznakom za strukturu, nakon što je bila dostupna informacija o jeziku. Informacije o jeziku i strukturi odvojene su tabulatorom, jer se oznaka za tabulator ne nalazi u korpusu. Kako izgleda jedna natuknica označena na obje razine, može se vidjeti na primjeru *Acetire*:

| | | |
|-----------|----|----------|
| Acetire | it | formlem |
| , | it | pc |
| ed | it | hint |
| inacetire | it | hint |
| , | it | pc |
| farfi | it | hint |
| aceto | it | hint |
| . | it | pc |
| <i> | la | cittrans |
| Acefco | la | cittrans |
| , | la | pc |
| ◇ | la | lb |
| fcis | la | v |
| . | la | pc |

| | | |
|----------------|----|------------|
| </i> | la | cittrans |
| Uşoftitifehr | | cittrans |
| , | hr | pc |
| oftimfe | hr | v |
| , | hr | pc |
| tiòsàmfe | hr | v |
| . | hr | pc |
| Uşce- chife | | hrcittrans |
| , | hr | pc |
| uşcişcèmfehr | | v |
| uşcegaòsàmfe | | hrv |
| . | hr | pc |
| Şcighnuti- fe | | hrcittrans |
| , | hr | pc |
| ghgnivàmfehr | | v |
| , | hr | pc |
| şcighnuòsàmfe | | hrv |
| . | hr | pc |
| Uşgljuti- tife | | hrcittrans |
| , | hr | pc |
| gljutimfe | hr | v |
| , | hr | pc |
| gljutiòsàmfe | | hrv |
| . | hr | pc |
| Proşuknutihr | | cittrans |
| , | hr | pc |
| ◊ | hr | lb |
| gnivam | hr | v |

| | | |
|-----------------|----|------------|
| , | hr | pc |
| proşuknuòfam | | hrv |
| . | hr | pc |
| Pochi | hr | cittrans |
| na | hr | cittrans |
| ozat | hr | cittrans |
| . | hr | pc |
| <pb> | hr | pb |
| Pochi | hr | cittrans |
| na | hr | cittrans |
| şcizzu | hr | cittrans |
| , | hr | pc |
| fi | it | cittrans |
| dice | it | cittrans |
| del | it | cittrans |
| vino | it | cittrans |
| . | it | pc |
| Ace- tire | it | cittrans |
| del | it | cittrans |
| pane | it | cittrans |
| . | hr | pc |
| Uşkvaſitife | | hrcittrans |
| , | hr | pc |
| fivàmfe | hr | v |
| , | hr | pc |
| uşkva- fiòsàmfe | | hrv |
| . | hr | pc |
| Prikifnutihr | | cittrans |

| | | |
|----------------|----|------------|
| , | hr | pc |
| fgnivam | hr | v |
| , | hr | pc |
| prikifnuò- fam | | hrv |
| . | hr | pc |
| Acetirfi | it | cittrans |
| del | it | cittrans |
| latte | it | cittrans |
| . | it | pc |
| Uşvifctàtfe | | hrcittrans |
| , | hr | pc |
| fctu- jèmfe | | hrv |
| , | hr | pc |
| uşvifctaòsàmfe | | hrv |
| . | hr | pc |
| Proşuknutihr | | cittrans |
| , | hr | pc |
| gni- vam | hr | v |
| , | hr | pc |
| proşuknuòfam | | hrv |
| . | hr | pc |
| v. | it | abbr |
| g. | it | abbr |
| Mljekòje | hr | cittrans |
| pro- şuknulo | | hrcittrans |
| . | hr | pc |
| Uşkisjelìtfe | | hrcittrans |
| , | hr | pc |

| | | |
|-------------------|----|----------|
| livàmfe | hr | v |
| , | hr | pc |
| uškifelilòsàm- fe | | hrv |
| . | hr | pc |
| Far | it | cittrans |
| acetire | it | cittrans |
| . | it | pc |
| <i> | la | cittrans |
| In | la | cittrans |
| acetum | la | cittrans |
| converterela | | cittrans |
| . | la | pc |
| </i> | la | cittrans |
| Ci- niti | hr | cittrans |
| uşoftiti | hr | cittrans |
| . | hr | pc |
| Ciniti | hr | cittrans |
| , | hr | pc |
| dàfe | hr | cittrans |
| uşofti | hr | cittrans |
| . | hr | pc |
| Uşkvafiti | hr | cittrans |
| ◊ | hr | lb |
| kruh | hr | cittrans |
| ; | hr | pc |
| far | it | cittrans |
| acetir | it | cittrans |
| il | it | cittrans |

| | | |
|-------------|----|----------|
| pane | it | cittrans |
| . | it | pc |
| Ukisjelitih | hr | cittrans |
| , | hr | pc |
| far | it | cittrans |
| ◊ | it | lb |
| acetir | it | cittrans |
| il | it | cittrans |
| latte | it | cittrans |
| . | it | pc |
| L' | it | cittrans |
| acetire | it | cittrans |
| , | it | pc |
| il | it | cittrans |
| farfi | it | cittrans |
| aceto | it | cittrans |
| . | it | pc |
| ◊ | it | lb |
| <i> | la | cittrans |
| In | la | cittrans |
| acetum | la | cittrans |
| mutatio | la | cittrans |
| . | la | pc |
| </i> | la | cittrans |
| Uşoftegne | hr | cittrans |
| , | hr | pc |
| egna | hr | gensg |
| . | hr | pc |

| | | |
|----------------|----|----------|
| n. | hr | pos |
| ◊ | hr | lb |
| Il | it | cittrans |
| farfi | it | cittrans |
| aceto | it | cittrans |
| del | it | cittrans |
| pane | it | cittrans |
| . | it | pc |
| Uškva fegne hr | | cittrans |
| , | hr | pc |
| gna | hr | gensg |
| . | hr | pc |
| n. | hr | pos |

Ovakav prikaz moguće je alternativno prikazati i kao „mentalnu mapu“ na temelju metodologije opisane u doktorskoj disertaciji pod nazivom „Automatsko stvaranje konceptualne mape iz nestrukturiranoga teksta na hrvatskome jeziku“.¹⁹⁰

¹⁹⁰Žubrinčić,

Krunoslav.

"Automatsko stvaranje konceptualne mape iz nestrukturiranoga teksta na hrvatskome jeziku". Doktorska disertacija, Fakultet elektrotehnike i računarstva, Sveučilište u Zagrebu, 2014.

6.3.3. Metoda

U istraživanju je korišten najsuvislji algoritam nadziranog strojnog učenja za označavanje sekvenci nazvan uvjetna nasumična polja (engl. *conditional random fields, CRF*).¹⁹¹

Uvjetna nasumična polja (CRF) su statistička metoda za predviđanje strukture, koja ima mogućnost predvidjeti više međusobno zavisnih varijabli. Primjenjuje se u različitim područjima poput računalnog vida, bioinformatike te, naravno, računalne obrade prirodnoga jezika.¹⁹²

Alat korišten za učenje i evaluaciju je *crfsuite*¹⁹³, koji implementira metodu uvjetnih nasumičnih polja. Alat implementira više različitih vrhunskih metoda strojnog učenja, a za ovo istraživanje je korišten pasivno agresivni algoritam za učenje jer je pokazao najbolje rezultate. Alat omogućava brzo učenje i označavanje podataka, jednostavan format podataka te dizajniranje proizvoljnog broja značajki za svaku jedinicu. Također je moguće ovim alatom izračunati rezultate modela evaluiranih na skupu za provjeru (preciznost, potpunost, F mjera).

Problemu označavanja jezika pojavnice i problemu označavanja strukture pristupilo se jednako. Prvo su definirani potencijalno zanimljivi skupovi značajki koji bi mogli dati bolje rezultate od samih podataka. Zatim je izvršeno

¹⁹¹Lafferty, J., McCallum, A., Pereira, F. (2001). "Conditional random fields: Probabilistic models for segmenting and labeling sequence data". Proc. 18th International Conf. on Machine Learning. Morgan Kaufmann. pp. 282-289.

¹⁹²Prema članku Sutton, Charles; McCallum, Andrew. An Introduction to Conditional Random Fields. // Foundations and Trends in Machine Learning. 4, 4(2011), str. 267-373. <http://homepages.inf.ed.ac.uk/csutton/publications/crftut-fnt.pdf> (7.9. 2014.).

¹⁹³<http://www.chokkan.org/software/crfsuite/>

mjerjenje uspješnosti skupova značajki. Konačno, objedinjeni su svi korisni skupovi značajki kako bi se postigao najbolji mogući rezultat s definiranim značajkama.

6.3.3.1 *Evaluacijske mjere*

Uobičajene mjere korištene za evaluaciju modela u području obrade prirodnog jezika su preciznost, potpunost i F mjera. Preciznost mjeri koliko je dohvaćenih slučajeva relevantno, dok potpunost mjeri koliko je relevantnih slučajeva dohvaćeno. Kako bi objasnili način računanja navedenih mjera, moramo uvesti četiri izraza: istinito pozitivno rješenje (IP), istinito negativno rješenje (IN), lažno pozitivno rješenje (LP) i lažno negativno rješenje (LN). Izrazi „pozitivno“ i „negativno“ odnose se na predviđanja klasifikatora, dok se izrazi „istinit“ i „lažan“ odnosi na to je li predviđanje u skladu s promatranom varijablom.¹⁹⁴

Preciznost¹⁹⁵ se može definirati na sljedeći način:

$$PR = \frac{IP}{IP + LP}$$

odnosno kao odnos točno pronađenih slučajeva i ukupno pronađenih slučajeva, dok je definicija potpunosti¹⁹⁶ sljedeća:

¹⁹⁴Precision and recall // Wikipedia, The Free Encyclopedia.
http://en.wikipedia.org/wiki/Precision_and_recall (7. 9. 2014.).

¹⁹⁵Ljubešić, Nikola. “Pronalaženje događaja u višestrukim izvorima informacija.” Doktorska disertacija. Filozofski fakultet, Sveučilište u Zagrebu, 2009.

¹⁹⁶Ibid.

$$PO = \frac{IP}{IP + LN}$$

odnosno kao odnos točno pronađenih slučajeva i svih točnih slučajeva.

F mjera je harmonična sredina između preciznosti i potpunosti, klasična mjera u području obrade prirodnog jezika. F mjera koristi faktor β . U slučaju da je $\beta > 1$, više naglašava potpunost, dok u slučaju da je $\beta < 1$, više naglašava preciznost. Računa se na sljedeći način¹⁹⁷:

$$F_{\beta} = \frac{(\beta^2 + 1) * PR * PO}{\beta^2 * PR + PO}$$

Osim prethodno navedene tri mjere, točnost¹⁹⁸ je dodatna mjera koju smo koristili za evaluaciju, a njezina definicija je sljedeća:

$$točnost = \frac{IP + IN}{IP + IN + LP + LN}$$

odnosno omjer svih točnih slučajeva i svih mogućih slučajeva.

6.3.3.2 Faze istraživanja

Istraživanje je provedeno u tri faze. Prva faza sastoji se od provjere najočitije značajke, odnosno utječeli način pisanja natuknice na rezultat: izvorni oblik pojavnice i njezin oblik pretvoren u kurent (pisano sve malim slovima). Očekuje se da će jedan od ta dva načina pisanja natuknice ostvariti bolji rezultat.

¹⁹⁷Ibid.

¹⁹⁸Precision and recall // Wikipedia, The Free Encyclopedia.
http://en.wikipedia.org/wiki/Precision_and_recall (7.9. 2014.).

Stoga se značajka s ostvarenim boljim rezultatom upotrebljava u sljedećim fazama istraživanja kao osnovnaznačajka.

U drugoj fazi istraživanja provjerava se utjecaj dodatnih značajki na rezultatestrojnog učenja. Kao osnovnaznačajka koristi se ona iz prve faze. Kod jezičnih oznaka kao dodatnaznačajkamjerena je logička varijabla kojom je zabilježen podatak je li originalna pojava pisana kurentom ili nije. Zatim je mjerena frekvencija pojavljivanja određenog troslova. Nadalje je provjeren utjecaj prethodnih i sljedećih pojava, gdje se N kreće od 1 do 3. Konačnaznačajkamjerena u eksperimentu je logička varijabla kojom je zabilježen podatak je li prethodna ili sljedeća pojava pisana kurentom ili nije.

Kod strukturnih oznaka kao dodatnaznačajkamjerena je logička varijabla kojom je zabilježen podatak je li originalna pojava pisana kurentom ili nije. Zatim je provjeren utjecaj prethodnih i sljedećih pojava, gdje se N kreće od 1 do 4. Nadalje, značajka upotrijebljena u eksperimentu je logička varijabla kojom je zabilježen podatak je li prethodna ili sljedećapojava pisana kurentom ili nije. Budući da u ovom skupu podataka svaka pojava sadrži podatak o jeziku, provjeren je utjecaj teznačajke na rezultat. Konačno, provjeren je utjecaj sufiksa na rezultat, gdje su uzeta posljednja četiri znaka pojavnice.

U trećoj fazi istraživanja, sveznačajke koje pokazuju bolje rezultate naspram rezultata postignutog osnovnom značajkomobjedinjenesu u jedan eksperiment. Time se postiže najbolji rezultat s definiranimznačajkama.

U budućim istraživanjima moguće je unijeti dodatneznačajke koje bi dale bolje rezultate od navedenih, te time dobiti najpovoljniji skup značajki.

6.3.3.3 *n*-struka unakrsna validacija

Svaki parametar evaluiran je računajući točnost putem 10-struke unakrsne validacije (*10-fold cross-validation*). U navedenoj metodi se podskupovi za učenje slučajnim uzorkovanjem grade iz skupa za učenje.

Unakrsna validacija je metoda validacije koja omogućava procjenu kako će se rezultati statističke analize upćiti na nezavisnom skupu podataka. Uglavnom se koristi u slučajevima kada se želi procijeniti koliko će točan biti prediktivni model nad nepoznatim podacima. Prediktivni modeli uglavnom sadrže dva skupa podataka. Jedan skup tih podataka su podaci za učenje, koji sadrži poznate podatke, dok drugi skup sadrži nepoznate podatke. Cilj unakrsne validacije je definiranje skupa podataka za provjeru modela u fazi učenja, čime se stvara novi podskup podataka za učenje, odnosno skup za validaciju. Cilj unakrsne validacije je nastojanje da se izbjegne preučenost modela te da se vidi kako će se model upćiti na nepoznatim podacima.¹⁹⁹

Unakrsna validacija upotrijebljena u ovom istraživanju je takozvana *n*-struka unakrsna validacija, gdje *n* iznosi 10. Kod ove vrste unakrsne validacije izvorni uzorak je nasumično podijeljen na *n* (odnosno 10) podskupova jednakih veličina. Jedan podskup koristi se kao skup za validaciju za testiranje modela, dok se ostalih *n*-1 (odnosno 9) podskupova koristi kao skupovi za učenje. Zatim se proces unakrsne validacije ponavlja *n* puta, gdje se svaki od *n* podskupova koristi točno jednom kao podaci za validaciju. Dobiveni rezultati su uprosječeni kako bi se dobila jedna procjena.²⁰⁰

¹⁹⁹ Cross-validation (statistics) // Wikipedia, The Free Encyclopedia. [http://en.wikipedia.org/wiki/Cross-validation %28statistics%29](http://en.wikipedia.org/wiki/Cross-validation_%28statistics%29) (7.9. 2014.).

²⁰⁰Ibid.

U sljedećim odlomcima opisani su eksperimenti provedeni u ovom istraživanju: označavanje jezika (6.3.4.) i strukture (6.3.5.) rječnika, eksperiment s ručnim označavanjem i brzinom ispravaka (6.3.6.) te konačan eksperiment (6.3.7.). Na kraju je prikazana i analizirana krivulja učenja (6.3.8.).

6.3.4. Označavanje jezika

U prvoj fazi istraživanja provedeno je označavanje jezika sa skupom od tri oznake. Istraživanje je provedeno sa sljedećim značajkama:

- `pojavnica` – pojavnica u originalnom obliku
- `mpojavnica` – pojavnica pretvorena u kurent (mala slova)
- `mslovabool` – logička (*Boolean*) varijabla koja poprima dvije vrijednosti: je li originalna pojavnica pisana kurentom ili nije
- `troslov` – frekvencija određenog troslova (*trigraph*)
- `preNpojavnica` i `postNpojavnica` – prethodna i sljedeća pojavnica (N=1..3)
- `preNmslovabool` i `postNmslovabool` – logička (*Boolean*) varijabla koja poprima dvije vrijednosti: je li prethodna ili sljedeća pojavnica pisana kurentom.

Slijedi primjer vrijednosti značajki za četvrtu pojavnicu u sljedećem nizu:

```
radici
.
V.
Barbare
.
```

Radicare

.

Vrijednosti značajki gore navedenih pojava su sljedeće:

```
pojavnica=Barbare
mpojavnica=barbare
mslovabool=False
troslov=_ba:1
troslov=bar:2
troslov=arb:1
troslov=rba:1
troslov=are:1
troslov=re_:1
pre1pojavnica=V.
pre2pojavnica=.
pre3pojavnica=radici
post1pojavnica=.
post2pojavnica=radicare
post3pojavnica=.
pre1mslovabool=False
pre2mslovabool=True
pre3mslovabool=True
post1mslovabool=True
post2mslovabool=True
post3mslovabool=True
```

Postignuta točnost određenim značajkama nalazi se u Tablici 6.3. Budući da su pojavnice pretvorene u kurent (mala slova) postigle bolje rezultate nego pojavnice u originalnom obliku, preostali eksperimenti koriste pojavnice pretvorene u kurent kao osnovnu značajku.

Tablica 6.3 Točnost označavanja jezika različitim značajkama

| značajke | točnost |
|---|----------------|
| pojavnica | 0,93224 |
| mpojavnica | 0,94143 |
| mpojavnicamslovabool | 0,95405 |
| mpojavnicatroslov | 0,97107 |
| mpojavnicapreNmpojavnicapostNmpojavnica
N=1 | 0,97188 |
| mpojavnica preNmpojavnicapostNmpojavnica
N=1..2 | 0,97997 |
| mpojavnica preNmpojavnicapostNmpojavnica
N=1..3 | 0,97697 |
| mpojavnica preNmslovabool
postNmslovabool N=1 | 0,94475 |
| mpojavnica preNmslovabool
postNmslovabool N=1..2 | 0,95086 |
| mpojavnica preNmslovabool | 0,94142 |

| značajke | točnost |
|--|----------------|
| postNmslovabool N=1..3 | |
| mpojavnicamslovabool troslov
preNpojavnicapostNpojavnica
preNmslovabool postNmslovabool N=1..2 | 0,98413 |

Najinformativnija značajka su troslovi pojavaice te prethodne i sljedeće pojavaice. Korištenjem dviju prethodnih i sljedećih pojavaica postižu se nešto bolji rezultati nego korištenjem jedne ili triju prethodnih i sljedećih pojavaica. Iz tog su razloga u zadnjem parametru objedinjene sveznačajke koje daju najbolje rezultate:

- mpojavnica – pojavaice pretvorene u kurent (mala slova)
- troslov – frekvencija određenog troslova (*trigraph*)
- mslovabool – logička (*Boolean*) varijabla koja poprima dvije vrijednosti: je li originalna pojavaica pisana kurentom ili nije
- preNmslovabool i postNmslovabool – logička (*Boolean*) varijabla koja poprima dvije vrijednosti: je li N pojavaica s lijeva ili s desna pisana kurentom
- preNpojavnica i postNpojavnica – dvije prethodne i sljedeće pojavaice.

Ovaj skup značajki postiže najbolje rezultate, i pri tome ostvaruje za 1 % veću točnost nego korištenjem pojavaice pretvorene u kurent i okolinom od dvije pojavaice s lijeva i s desna.

U Tablici 6.4 navodi se preciznost, potpunost i F_1 mjera jezičnog klasifikatora za određene kategorije. Klasifikator postiže najbolje rezultate za

latinski jezik za sve tri navedene mjere: preciznost (P) iznosi 0,99815, potpunost (R) 0,99938 te F₁ mjera (F1) 0,99878. Budući da je latinski dio rječničke natuknice uvijek omeđen posebnim oznakama, takav rezultat se i očekivao. Klasifikator postiže bolji rezultat preciznosti za talijanski jezik (0,9829) nego za hrvatski jezik (0,97953). Razlog tome mogao bi ležati u činjenici da je početak rječničke natuknice uvijek na talijanskom jeziku. Bolji rezultat potpunosti postiže se za hrvatski (0,99067) nego za talijanski jezik (0,95831), što bi se moglo objasniti podatkom da je preko polovine (53 %) pojavnica označeno hrvatskim, a tek nešto više od četvrtine (26 %) talijanskim jezikom.

Tablica 6.4 Uspješnost konačnog jezičnog klasifikatora

| lang | match | model | ref | P | R | F1 |
|------|--------|--------|--------|----------------|----------------|----------------|
| hr | 4354,0 | 4445,0 | 4395,0 | 0,97953 | 0,99067 | 0,98507 |
| it | 1839,0 | 1871,0 | 1919,0 | 0,9829 | 0,95831 | 0,97045 |
| la | 1618,0 | 1621,0 | 1619,0 | 0,99815 | 0,99938 | 0,99876 |

6.3.5. Označavanje strukture

U drugoj fazi istraživanja provedeno je označavanje struktura rječničkih natuknica, odnosno skupom koji ima 19 oznaka. Istraživanje provedeno nad strukturnim oznakama prati jednaku metodologiju kao i kod jezičnih oznaka. Korištene su sljedeće značajke:

- pojavnica – pojavnica u originalnom obliku
- mpojavnica – pojavnica pretvorena u kurent (mala slova)

- `mslovabool` – logička (*Boolean*) varijabla koja poprima dvije vrijednosti: je li originalna pojava pisana kurentom ili nije
- `preNpojavnica` i `postNpojavnica` – prethodna i sljedeća pojava (N=1..4)
- `preNmslovabool` i `postNmslovabool` – logička (*Boolean*) varijabla koja poprima dvije vrijednosti: je li prethodna i sljedeća pojava pisana kurentom (N=1..4)
- `jezik` – jezik pojavnice
- `nastavakN` – nastavak pojavnice duljine N=4.

Postignuta točnost određenim značajkama prikazana je u Tablici 6.5. Prethodne i sljedeće četiri pojavnice pojedinačnu značajku koja pokazuje najbolje rezultate. Međutim, kada objedinimo sve značajke s najboljim rezultatima, točnost se povećava za gotovo 2 %. Tesu značajke sljedeće:

- `pojavnica` – pojava u originalnom obliku
- `mslovabool` – logička (*Boolean*) varijabla koja poprima dvije vrijednosti: je li originalna pojava pisana kurentom ili nije
- `preNpojavnica` i `postNpojavnica` – četiri prethodne i sljedeće pojavnice
- `preNmslovabool` i `postNmslovabool` – logička (*Boolean*) varijabla koja poprima dvije vrijednosti: jesu li četiri prethodne i sljedeće pojavnice pisane kurentom
- `jezik` – jezik pojavnice
- `nastavakN` – nastavak pojavnice duljine N=4.

Tablica 6.5 Točnost označavanja strukture različitim značkama

| značajke | točnost |
|--|----------------|
| pojavnica | 0,85993 |
| mpojavnica | 0,85538 |
| pojavnicamslovabool | 0,8934 |
| pojavnica preNpojavnicapostNpojavnica
N=1 | 0,90388 |
| pojavnica preNpojavnicapostNpojavnica
N=1..2 | 0,93794 |
| pojavnica preNpojavnicapostNpojavnica
N=1..3 | 0,94994 |
| pojavnica preNpojavnicapostNpojavnica
N=1..4 | 0,94219 |
| pojavnica preNmslovabool
postNmslovabool N=1 | 0,87586 |
| pojavnica preNmslovabool
postNmslovabool N=1..2 | 0,87706 |
| pojavnica preNmslovabool
postNmslovabool N=1..3 | 0,88755 |
| pojavnica preNmslovabool
postNmslovabool N=1..4 | 0,89588 |
| pojavnicajezik | 0,86555 |
| pojavnicanastavakN N=4 | 0,87192 |

tekstu, dok se ispred interpunkcija uvijek nalazi razmak, osim ako točka nije sastavni dio kratice. Treći najbolji rezultat (0,99087) klasifikator postiže za oznaku *v*, a razlog tome mogla bi biti činjenica da se oznaka odnosi na glagolske nastavke, koji se redovito ponavljaju u jednakom obliku. Najlošije rezultate potpunosti klasifikator postiže za oznake *genpl* i *adj*, kao i kod preciznosti, zato što su oznake rijetke u ručno označenom korpusu. Iznenadjujući je treći najlošiji rezultat (0,57143) potpunosti, a odnosi se na oznaku *cb*. Ta je oznaka izričito označena u rječničkom tekstu na dva načina: može biti samostalna, ali se može nalaziti i unutar jedne riječi, gdje je ostavljena kao sastavni dio pojavnice. Možemo pretpostaviti da samostalno označen prelazak u novi stupac omogućava potpunu preciznost. Međutim, očito klasifikator nailazi na prepreke kada je oznaka za prelazak u novi stupac sastavni dio pojavnice.

Najbolja tri rezultata F_1 mjere postignuti su za oznaku *lb* (1,0), oznaku *pc* (0,99786) i oznaku *v* (0,9819). Iz navedenoga je vidljivo da klasifikator postiže 100 %-tne rezultate svih triju mjera za oznaku *lb*, dok za oznaku *pc* postiže za sve tri mjere jedan od tri najbolja rezultata. S druge strane, najgori rezultat postignut je za oznake *adj* (0,0), *genpl* (0,0) i *pb* (0,6), što je i očekivano, jer se te oznake u ručno označenom korpusu rijetko pojavljuju.

Tablica 6.6 Uspješnost konačnog strukturnog klasifikatora

| struct | match | model | ref | P | R | F1 |
|---------------|--------------|--------------|------------|----------|----------|-----------|
| abbr | 36,0 | 42,0 | 46,0 | 0,85714 | 0,78261 | 0,81818 |
| adj | 0,0 | 0,0 | 2,0 | 0,0 | 0,0 | 0,0 |
| adjf | 104,0 | 107,0 | 105,0 | 0,97196 | 0,99048 | 0,98113 |
| adjn | 105,0 | 111,0 | 114,0 | 0,94595 | 0,92105 | 0,93333 |

| struct | match | model | ref | P | R | F1 |
|---------------|--------------|--------------|------------|---------------|----------------|----------------|
| bibl | 78,0 | 82,0 | 79,0 | 0,95122 | 0,98734 | 0,96894 |
| cb | 4,0 | 4,0 | 7,0 | 1,0 | 0,57143 | 0,72727 |
| citex | 591,0 | 619,0 | 620,0 | 0.95477 | 0,95323 | 0,954 |
| cittrans | 2717,0 | 2776,0 | 2838,0 | 0,97875 | 0,95736 | 0,96794 |
| formlem | 100,0 | 103,0 | 111,0 | 0,97087 | 0,9009 | 0,93458 |
| genpl | 0,0 | 0,0 | 1,0 | 0,0 | 0,0 | 0,0 |
| gensg | 167,0 | 172,0 | 170,0 | 0,97093 | 0,98235 | 0,97661 |
| hint | 333,0 | 438,0 | 364,0 | 0,76027 | 0,91484 | 0,83042 |
| lb | 450,0 | 450,0 | 450,0 | 1,0 | 1,0 | 1,0 |
| pb | 3,0 | 5,0 | 5,0 | 0,6 | 0,6 | 0,6 |
| pc | 2100,0 | 2104,0 | 2105,0 | 0,9981 | 0,99762 | 0,99786 |
| pos | 180,0 | 185,0 | 183,0 | 0,97297 | 0,98361 | 0,97826 |
| ref | 31,0 | 32,0 | 34,0 | 0,96875 | 0,91176 | 0,93939 |
| v | 217,0 | 223,0 | 219,0 | 0,97309 | 0,99087 | 0,9819 |
| xr | 31,0 | 32,0 | 32,0 | 0,96875 | 0,96875 | 0,96875 |

6.3.6. Eksperiment s ručnim označavanjem i brzinom ispravaka

Sljedećim je eksperimentom odgovoreno na pitanje je li ispravljanje automatskih oznaka za jezik i strukturu vremenski brže naspram ručnog označavanja, te ako je odgovor potvrđan, koliko je brže. Eksperiment se sastojao

od dva dijela, a svaki je dio trajao 60 minuta. U prvom su se dijelu ručno označavale pojavnice, dok su se u drugom dijelu automatski označene pojavnice pregledavale i ispravljale. Ručno označavanje te pregledavanje i ispravljanje vršio je stručni označitelj koji dobro poznaje della Bellin rječnik, odnosno u ovom slučaju autorica disertacije. Rezultati eksperimenta dostupni su u Tablici 6.7.

Tablica 6.7 Rezultati brzine ručnog označavanja i ispravljanja automatski označenih oznaka u 60 minuta

| | broj
pojavnica | pojavnica u
minuti |
|--------------------------|---------------------------|-------------------------------|
| ručno označavanje | 741 | 12,35 |
| ispravljanje | 3 439 | 57,32 |

U prvom je dijelu označitelj ručno označavao pojavnice za jezik i strukturu u vremenu od 60 minuta. Nasumično je odabrana natuknica od koje je krenulo označavanje, koje se nastavilo redoslijedom pojavljivanja pojava u korpusu. U navedenom razdoblju označena je 741 pojava, odnosno unesene su sveukupno 1 482 oznake. To znači da se u minuti može označiti 12,35 pojava.

U drugom dijelu eksperimenta označitelj je unaprijed imao automatski označene pojavnice s jezičnim i strukturnim oznakama, koje je trebalo u vremenu od 60 minuta pregledati i ispraviti eventualne pogreške. Nasumično je

odabrana natuknica od koje je krenulo pregledavanje i ispravljanje, koje se nastavilo redosljedom pojavljivanja pojava u korpusu. U navedenom razdoblju označeno je 3 439 pojava, odnosno pregledano je i po potrebi ispravljeno 6 878 oznaka. To znači da se može pregledati i označiti 57,32 pojava u minuti, odnosno da je ova metoda 4,64 puta brža od ručnog unošenja oznaka.

Iz eksperimenta se može zaključiti da se automatsko označavanje isplati s obzirom na činjenicu da je 4,64 puta brže od ručnog unošenja oznaka. Dodatna vrijednost ovog eksperimenta jest 4 180 pojava koje su naknadno ručno pregledane.

Ovim eksperimentom dodatno je ručno pregledano 4 180 pojava, čime je skup za učenje narastao na 12 520 pojava, tako da sada ručno označeni korpus čini 3,11 %, a skup neoznačenih podataka sadrži 390 608 odnosno 96,89 %. Ispravljanjem materijala možemo povećavati skup za učenje čime bi se dobili točniji modeli, a iz toga proizlazi da bi se time još više smanjilo vrijeme ispravljanja.

6.3.7. Završni eksperiment na skupu za provjeru

Kako bi detaljnije analizirali uspješnost klasifikatora, uobičajeno je prikazati tablicu zabuna. Tablica zabuna naziva se još i matrica zabuna (*confusion matrix*), matrica pogrešaka (*error matrix*) i kontigencijska tablica u statistici (*contingency table*). Tablica zabuna omogućava vizualizaciju rezultata algoritma korištenog za klasifikaciju. Redovi predstavljaju stvarne kategorije,

dok stupci predstavljaju predviđene kategorije. Na taj način možemo vidjeti koje su to kategorije koje algoritam često krivo označava.²⁰¹

Iz eksperimenta s ručnim označavanjem izdvojen je skup za provjeru veličine 7 987 pojava. Točnost klasifikatora za jezične oznake iznosi 0,97308. Tablica zabunaprikazana je u Tablici 6.8. Iz tablice je vidljivo da klasifikator ima najmanje problema s prepoznavanjem latinskog jezika. Budući da je latinski dio rječničke natuknice uvijek omeđen posebnim oznakama, takav rezultat je očekivan. Klasifikator ima najviše problema s hrvatsko-talijanskim parom. Rječničke natuknice često ne prate strukturu trojezičnog rječnika, pa tako niz jezika koji se pojavljuju nije uvijek talijansko-latinsko-hrvatski. Latinski dio natuknice izričito je omeđen posebnim oznakama, što bi odlično odvajalo talijanski od hrvatskog dijela. Međutim, ako se unutar natuknice nalazi složenica, često nedostaje latinski prijevod, čime se talijanski i hrvatski tada nalaze jedan pored drugog. Osim toga, za hrvatski jezik tada još nije postojao standardni pravopis, pa se della Bella oslanjao na praksu koja je postojala u talijanskom jeziku za bilježenje hrvatskih glasova. Sve su to mogući razlozi zašto klasifikator ima najviše problema s talijansko-hrvatskim parom.

Tablica 6.8 Tablica zabuna za jezične oznake

| | hr | it | la |
|----------------|-----------|-----------|-----------|
| hr | 5 128 | 19 | 1 |
| it | 194 | 1351 | 1 |
| la | 0 | 0 | 1 293 |
| 0,973081257043 | | | |

²⁰¹Confusion matrix.// Wikipedia, The Free Encyclopedia.
http://en.wikipedia.org/wiki/Confusion_matrix (8.9. 2014.).

Točnost klasifikatora za strukturne oznake iznosi 0,954801. Tablica zabuna nalazi se u Tablici 6.9. Iz tablice je vidljivo da klasifikator ima najviše problema s oznakom `citttrans`, koju zamjenjuje najčešće s `hint`, `citex` i `v`. Razlog tome može biti u sadržaju tih struktura natuknica, koje često sadrže slobodan tekst. Najbolje rezultate klasifikator postiže za oznaku `xr` koja je u svim slučajevima točno klasificirana te za oznake `lbi` `bi` `bl` koje je samo jednom pogrešno kategorizirao. Te su oznake u rječničkom tekstu izričito označene, stoga ne iznenađuje da one dobivaju najbolje rezultate.

Tri se oznake uopće nisu pojavile u skupu za provjeru, a to su `cb`, `pb` i `adj`.

Tablica 6.9 Tablica zabuna za strukturne oznake

| | cittrans | ref | lb | bibl | hint | cb | v | pos | pb | pc | abbr | citex | adjn | xr | gensg | formlem | adj | adjf |
|-----------------|-----------------|------------|-----------|-------------|-------------|-----------|----------|------------|-----------|-----------|-------------|--------------|-------------|-----------|--------------|----------------|------------|-------------|
| cittrans | 2 129 | 0 | 0 | 0 | 18 | 0 | 3 | 0 | 0 | 4 | 0 | 19 | 2 | 0 | 1 | 0 | 0 | 3 |
| ref | 17 | 43 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| lb | 0 | 0 | 506 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| bibl | 2 | 0 | 0 | 80 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| hint | 70 | 0 | 0 | 0 | 173 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 2 | 0 | 2 | 6 | 0 | 0 |
| cb | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| v | 32 | 0 | 0 | 0 | 5 | 0 | 568 | 0 | 0 | 0 | 0 | 0 | 19 | 0 | 1 | 0 | 0 | 2 |
| pos | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 253 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| pb | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| pc | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 618 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| abbr | 2 | 0 | 0 | 0 | 4 | 0 | 0 | 2 | 0 | 2 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| citex | 48 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 556 | 0 | 0 | 1 | 0 | 0 | 0 |
| adjn | 2 | 0 | 0 | 0 | 2 | 0 | 1 | 10 | 0 | 0 | 0 | 0 | 124 | 0 | 2 | 0 | 0 | 1 |
| xr | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 44 | 0 | 0 | 0 | 0 |
| gensg | 5 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 245 | 0 | 0 | 0 |
| formlem | 2 | 0 | 0 | 0 | 13 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 135 | 0 | 0 |
| adj | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| adjf | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 6 | 0 | 0 | 133 |
| 0,954801552523 | | | | | | | | | | | | | | | | | | |

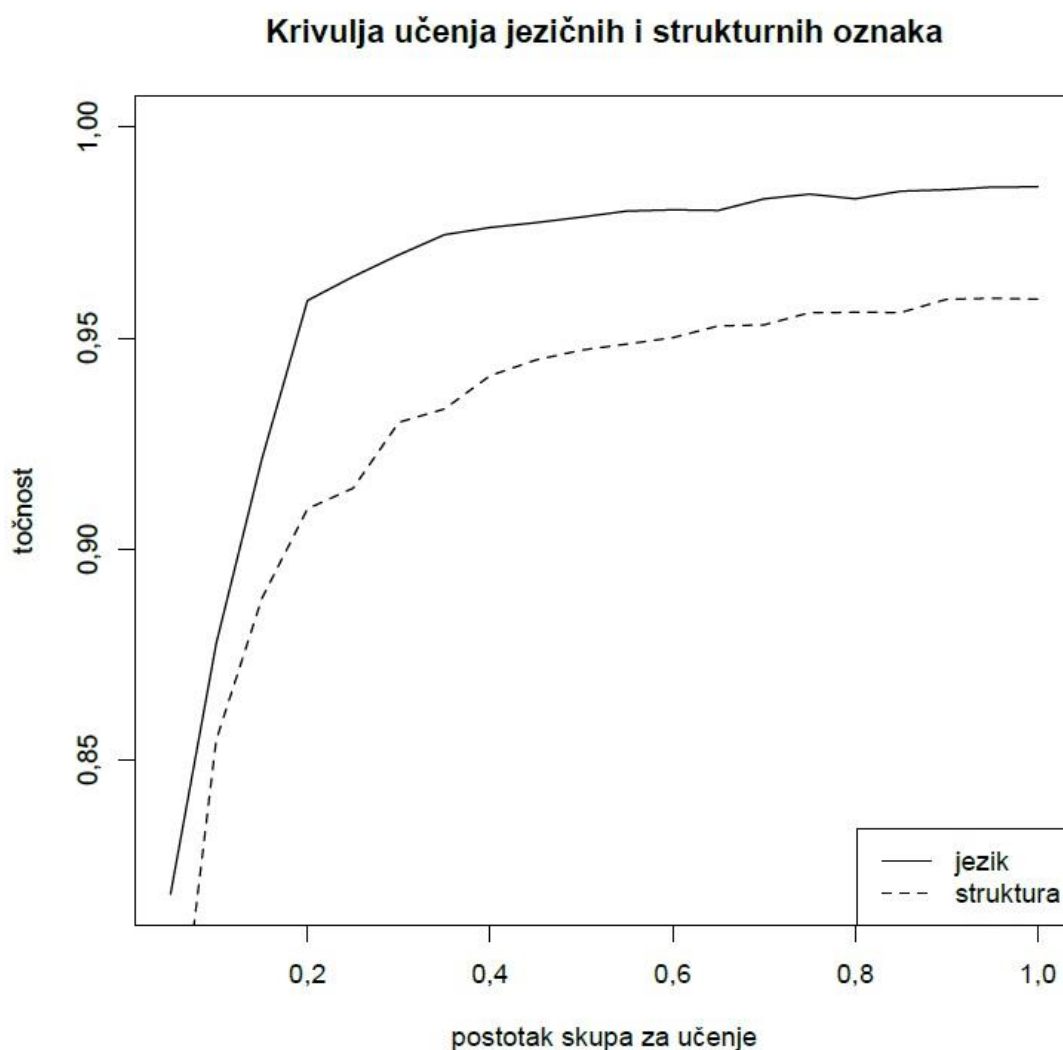
6.3.8. Krivulja učenja

U sljedećem dijelu rada prikazana je i analizirana krivulja učenja algoritma upotrijebljenog za strukturiranje podataka pomoću jezičnih i strukturnih oznaka. S obzirom na to da jezična razina sadrži samo tri oznake, dok strukturna sadrži 19 oznaka, očekuju se bolji rezultati za prvu skupinu, a nešto lošiji za drugu. Također se očekuje da će za jezičnu razinu biti potrebna manja količina podataka kako bi algoritam naučio većinu pravila, dok se za strukturnu razinu očekuje potreba za većom količinom podataka.

Krivulje učenja za jezičnu i strukturnu razinu prikazane su na Slici 6.2, gdje je vidljivo da algoritam bolje raspoznaje jezik nego strukturu. Krivulja jezičnih oznaka brže raste te je već nakon 20 % podataka (odnosno 1 668 pojavnica) postignuta točnost od gotovo 96 %. Na cjelokupnom skupu za provjeru postignuta je točnost od 98,59 %. Takav rezultat smatra se odličnim, s obzirom na činjenicu da se radi o tekstu iz 1785. godine kada pravopis hrvatskog jezika nije bio standardiziran, pa su bili česti neujednačeni načini zapisivanja hrvatskih glasova unutar istog teksta, u ovom slučaju rječnika. Razumljivo je da će neujednačenost zapisivanja u hrvatskom imati utjecaj na rezultate, jer više od polovine (točnije 53 %) pojavnica u ručno označenom korpusu čine pojavnice označene oznakom za hrvatski jezik.

Strukturne oznake postižu slabije rezultate te je učenje sporije. Potrebno je otprilike 40 % podataka (odnosno 3 336 pojavnica) kako bi se postigla točnost veća od 94 %, a konačan rezultat na cjelokupnom skupu za provjeru iznosi 95,92 %. Budući da strukturnih oznaka ima 19, takav je rezultat premašio očekivanja.

Slika 6.2 Krivulja učenja jezičnih i strukturnih oznaka



Iz podataka je očito da jezična razina postiže bolje rezultate od strukturne razine (98,59 % naspram 95,92 %). Osim toga, za jezičnu razinu većina je učenja obavljena na uzorku od 20 %, dok je za strukturnu razinu bila potrebna dvostruko veća količina podataka, odnosno uzorak od 40 %.

Budući da krivulje za obje razine još značajno rastu, moguće je popraviti izlazne rezultate trenutnog algoritma, tako da se uključe dodatni i novi podaci. Time bi se dodatno povećala točnost te smanjilo vrijeme potrebno za naknadnu ručnu obradu rječnika. Drugim riječima, postojeći algoritam je korektan i može

ga se upotrebljavati s dovoljnom točnošću za automatsko označavanje strukture tekstova naših starih rječnika, odnosno može se reći da je algoritam isplativ.

6.4. Povezivanje koncepata na resurse

Zahvaljujući formaliziranom prikazu rječničkih natuknica putem TEI oznaka, moguće je mapirati elemente na LMF (*Lexical Markup Framework*).²⁰² LMF²⁰³ je ISO standard za označavanje leksikona u području prirodne obrade jezika i strojno čitljivih rječnika. Označavanje resursa dodatnim standardom proširuje potencijalne zainteresirane pojedince i projekte, koji bi se mogli nadovezati na resurs.

Osim uvođenja novog standarda, koristan bi bio sustav pretraživanja povijesnih dokumenata koji bi omogućio uparivanje upita na suvremenom jeziku s povijesnim oblicima pronađenim u dokumentima. Na taj bi način korisnici mogli postavljati upite na suvremenom jeziku i pravopisu.

Koncepte unutar rječnika moguće je prikazati putem RDF-a (*Resource Description Network*).²⁰⁴ Postoje pokušaji mapiranja LMF-a i RDF-a²⁰⁵ koje valja istražiti te vidjeti postoji li mogućnost automatskog mapiranja na RDF.

²⁰²Laurent Romary and Werner Wegstein, “Consistent Modeling of Heterogeneous Lexical Structures”, Journal of the Text Encoding Initiative [Online], Issue 3 | November 2012, Online since 15 October 2012.URL :<http://jtei.revues.org/540> ; DOI : 10.4000/jtei.540 (14. 4. 2014.).

²⁰³**Lexical Markup Framework** // Wikipedia, The Free Encyclopedia.
http://en.wikipedia.org/wiki/Lexical_Markup_Framework (5.9. 2014.).

²⁰⁴Lassila, Ora; Swick, Ralph R.M; World Wide Web Consortium.Resource Description Framework (RDF) Model and Syntax Specification, 1998.

²⁰⁵Sérasset, Gilles (2012). DBnary: Wiktionary as a LMF based Multilingual RDF network. *InLanguage Resources and Evaluation Conference, LREC 2012*, Istanbul.

Dodatni oblici formalizacije omogućili bi povezivanje koncepata na vanjske resurse, kao što su mrežno izdanje Hrvatske enciklopedije Leksikografskog zavoda²⁰⁶, Hrvatski jezični portal²⁰⁷, Linked Data^{208,209} i DBpedije^{210,211}.

6.5. Znanstveni doprinos modela

Znanstveni doprinos istraživanja je višestruk. S jedne strane, formaliziran je problem digitalizacije i obrade povijesnih rječničkih tekstova. S druge strane, razvijen je model višerazinskog prikaza tekstova starijih hrvatskih rječnika koji omogućava daljnja istraživanja ne samo u informacijskim znanostima, odnosno računalnoj leksikografiji kao dijelu informacijskih znanosti, nego i u ostalim lingvističkim i filološkim, ali i kulturološkim te povijesnim područjima i granama znanosti. Konačno, primijenjene su metode za automatsku i poluautomatsku obradu povijesnih tekstova.

Ovim radom provjerile su se dvije hipoteze:

²⁰⁶Enciklopedija Leksikografskog Zavoda Miroslav Krleža, Mrežno izdanje. <http://www.enciklopedija.hr/> (5. 9. 2014.).

²⁰⁷Hrvatski Jezični Portal.<http://hjp.novi-liber.hr/> (5. 9. 2014.).

²⁰⁸Linked Data - Connect Distributed Data Across the Web.” <http://linkeddata.org/> (5. 9. 2014.).

²⁰⁹Berners-Lee, Tim. “Linked Data - Design Issues,” July 27, 2006. <http://www.w3.org/DesignIssues/LinkedData.html> (5. 9. 2014.).

²¹⁰DBpedia.<http://dbpedia.org/About> (5. 9. 2014.).

²¹¹Bizer, Christian; Lehmann, Jens; Kobilarov, Georgi; Auer, Sören; Becker, Christian; Cyganiak, Richard; Hellmann, Sebastian. “DBpedia - A Crystallization Point for the Web of Data.” Web Semant. 7 (2009): 154-165.

1. S obzirom na stanje jezičnih tehnologija moguće je razviti metode za automatsku i poluautomatsku obradu povijesnih tekstova koji sadrže vremenski nezavisne inačice jezika.

Istraživanjem je dokazano da se suvremeni alati za strojno učenje i evaluaciju istih mogu uspješno primijeniti na povijesne tekstove. Povijesni tekstovi često sadrže više jezika u jednom dokumentu, što je i slučaj u ovom istraživanju. U istraživanju je korišten korpus koji sadrži hrvatski, talijanski i latinski jezik. Osim toga, u povijesnim tekstovima česti su neujednačeni načini zapisivanja određenih (odnosno hrvatskih) glasova unutar istog teksta, što statističkim metodama predstavlja prepreke jer dolazi do raspršenosti podataka. Usprkos tomu, suvremeni alat temeljen na statističkim metodama upotrijebljen u ovom istraživanju pokazao se uspješnim. Međutim, s obzirom na činjenicu da je istraživanje provedeno nad jednim izvorom, ne možemo ni potvrditi ni odbaciti drugi dio hipoteze o vremenski nezavisnim inačicama jezika, jer je dokument iz jednog razdoblja od jednog autora.

2. Zahvaljujući tim metodama, moguće je izgraditi sveobuhvatni model prikaza znanja u starijim hrvatskim rječnicima.

Predloženi model zadržava vjerodostojnost i izvornost prikaza svakog pojedinog rječnika, što je općenito bitno kod prikaza povijesnih tekstova. Osim toga, definirana je shema za označavanje digitaliziranih tekstova rječnika u skladu s najnovijim *de facto* standardima (*Text Encoding Initiative*, TEI), čime se dobiva na interoperabilnosti podataka. Označavanje strukture naporan je i dugotrajan posao, stoga je provedeno istraživanje za njeno automatsko označavanje. Korišten je alat kao i kod dokazivanja prve hipoteze, te se pokazao uspješnim u označavanju strukture rječničkog teksta. Osim toga, dokazano je da je ručno ispravljanje oznaka dodanih automatskom metodom znatno brže od ručnog unošenja oznaka.

Ovim radom dokazane su obje hipoteze postavljene na početku istraživanja, uz napomenu da je prva hipoteza samo u glavnini dokazana. Također su postignuti ciljevi istraživanja, postavljeni na početku.

6.6. Mogućnosti daljnjih istraživanja

Zahvaljujući formaliziranom prikazu rječničkih natuknica putem TEI oznaka, moguće je mapirati elemente na LMF (*Lexical Markup Framework*).²¹² LMF²¹³ je ISO standard za označavanje leksikona u području prirodne obrade jezika i strojno čitljivih rječnika. Označavanje resursa dodatnim standardom proširuje potencijalne zainteresirane pojedince i projekte, koji bi se mogli nadovezati na resurs.

Koncepte unutar rječnika moguće je prikazati putem RDF-a (*Resource Description Network*).²¹⁴ Postoje pokušaju mapiranja LMF-a i RDF-a²¹⁵ koje valja istražiti te vidjeti postoji li mogućnost automatskog mapiranja na RDF.

Dodatni oblici formalizacije omogućili bi povezivanje koncepata na vanjske resurse, kao što su mrežno izdanje Hrvatske enciklopedije Leksikografskog zavoda, Hrvatski jezični portal, Linked Data i DBpedije.

²¹²Laurent Romary and Werner Wegstein, “Consistent Modeling of Heterogeneous Lexical Structures”, Journal of the Text Encoding Initiative [Online], Issue 3 | November 2012, Online since 15 October 2012.URL :<http://jtei.revues.org/540> ; DOI : 10.4000/jtei.540 (15. 4. 2014.).

²¹³**Lexical Markup Framework** // Wikipedia, The Free Encyclopedia.
http://en.wikipedia.org/wiki/Lexical_Markup_Framework (5.9. 2014.).

²¹⁴Lassila, Ora; Swick, Ralph R.M; World Wide Web Consortium.Resource Description Framework (RDF) Model and Syntax Specification, 1998.

²¹⁵Sérasset, Gilles (2012). DBnary: Wiktionary as a LMF based Multilingual RDF network. *InLanguage Resources and Evaluation Conference, LREC 2012*, Istanbul.

Dodatno valja istražiti metode korištenja mnoštva (*Crowdsourcing*) u raznim fazama digitalizacije i obrade rječnika. Metodu je moguće iskoristiti prilikom ručnog ispravljanja transkribiranog teksta, označavanja natuknica TEI oznakama ili označavanjem skupa za učenje, čime bi se i rezultati automatskih metoda poboljšali.

Osim na razvoju samog modela, dodatna istraživanja bit će moguća upotrebom sustava kada on bude dostupan putem interneta. Vjerujemo da će sustav biti od interesa stručnjacima raznih područja i grana znanosti, kao što su lingvistička i filološka područja, ali i kulturološka i povijesna.

7. Zaključak

Razvoj informacijsko-komunikacijskih i jezičnih tehnologija omogućuje jednostavniji pristup i brži protok informacija, pohranu velikih količina podataka te sve jednostavnije rukovanje njima. Sve jednostavnija primjena tehnologija omogućuje njezino prodiranje u sve segmente ljudskog djelovanja, pa tako i u kulturnu djelatnost. Stoga ne začuđuje činjenica da se sve više projekata koji uključuju kulturnu baštinu okreću upravo navedenim tehnologijama kako bi omogućili pristup, protok, pohranu i rukovanje baštinom. Prvi korak koji je potrebno učiniti jest digitalizacija baštine, odnosno pretvorba analognih podataka odnosno dokumenata u digitalni oblik, kako bi se omogućila njihova računalna obrada. Razlozi pristupanju digitalizaciji baštine su raznoliki, a detaljno su razrađeni u poglavlju 5.1.1. Iako je za digitalizaciju prikladna različita vrsta građe (na primjer tekstovi, slike, zvuk, video...), uglavnom se teži pretvorbi tiskanih dokumenata u digitalne tekstove jer se oni mogu pretraživati i/ili automatski obrađivati. Upravo se digitalizacijom starijih hrvatskih tekstova bavi znanstveni projekt „Hrvatska rječnička baština i hrvatski europski identitet“ glavnog voditelja prof. dr. sc. Damira Borasa, u sklopu kojeg je izrađena ova disertacija.

Sve veća dostupnost povijesnih tekstova u digitalnom obliku ima za posljedicu i sve veće zanimanje za primjenu metoda, a time i alata, za obradu prirodnoga jezika na takvoj vrsti teksta.

Cilj istraživanja ovog rada je razvoj modela koji omogućava prikaz znanja u tekstovima starijih hrvatskih rječnika. Rječnici sadrže svome dobu suvremeni jezik koji se razvijao tijekom stoljeća, te su u svakome prikazana

različita dijakronijska stanja jezika, vremenski i pravopisno omeđena. Sedam starih hrvatskih rječnika obuhvaćenih ovim radom sadrže inačice hrvatskoga jezika u rasponu od preko 300 godina: od 1595. do 1901. Model zadržava vjerodostojnost i izvornost prikaza svakog pojedinog jezika. Model ujedno omogućava označavanje digitaliziranih tekstova rječnika u skladu s najnovijim *de facto* standardima (*Text Encoding Initiative*, TEI), koji omogućavaju interoperabilnost s drugim jezičnim resursima, alatima i sustavima za obradu povijesnih tekstova. Osim toga, primijenjene su metode za automatsku i poluautomatsku obradu digitaliziranih povijesnih tekstova, čime se ubrzao i pojednostavio proces njihove obrade.

Danas u svijetu postoji nebrojeno mnogo rječnika i enciklopedija u elektroničkom obliku, bilo u tekstualnom bilo u strukturiranom obliku rječničke (enciklopedijske) baze podataka s različitim indeksima, mogućnostima pretraživanja i slično. Suvremeni su elektronički rječnici uglavnom sustavi potpuno zatvoreni u sebe i ne omogućuju međusobno povezivanje niti kritičku usporedbu jer su pretežno trgovačke prirode, ali jednako i u slučajevima kad omogućuju slobodan pristup, taj pristup ograničen je sustavom kroz koji se pristupa podacima i koji ne omogućuje preuzimanje podataka.

Istovremeno, postoji veoma mnogo projekata kojima je cilj objavljivanje stare tekstualne građe, biblijskih tekstova, književnih djela, pa i rječnika, ali do sada za hrvatski jezik rječnici na taj način nisu obrađivani. Detaljna analiza određenog broja međunarodnih i domaćih projekata sličnih znanstvenom projektu „Hrvatska rječnička baština i hrvatski europski identitet“ može se naći u poglavlju 4.

Ipak, povijesni tekstovi imaju nekoliko jezičnih značajki koje predstavljaju prepreke za primjenu alata za obradu prirodnoga jezika u onom obliku u kakvom trenutno postoje. Naime, primjena algoritama učenih na jednoj

domeni daje slabije rezultate kada se primijene na tekstove drugih domena ili žanrova. Razlozi leže u razlikama na razini vokabulara, sintakse i semantike određenih područja i žanrova. Navedeni problemi mogu se preslikati i na cijelo područje obrade prirodnoga jezika, nevezano o tome radi li se o suvremenim ili povijesnim tekstovima.

Iako su povijesni jezici prirodni jezici, baš kao i jezici kojima danas govorimo i pišemo, postoje četiri temeljne značajke koje otežavaju primjenu alata za obradu prirodnoga jezika na povijesnim tekstovima: nedostatak standardne inačice, nedostatak standardnog pravopisa, nedostupnost tekstova u digitalnom obliku te nepostojeći resursi i alati za obradu prirodnog jezika za povijesne tekstove. Detaljnija analiza o obradi prirodnoga jezika i povijesnih tekstova nalazi se u poglavlju 5.

Stoga je potrebno razviti navedene metode, a time i alate, za povijesne tekstove, posebno tekstove na hrvatskom jeziku, na obradu kojih se i odnosi ovaj rad. Osim prethodno navedenih značajki povijesnih jezika, dodatne probleme predstavljaju pomaci u značenjima riječi te uporaba više od jednog jezika unutar jednog dokumenta. Također bi bilo potrebno omogućiti uparivanje upita na suvremenom jeziku s povijesnim oblicima pronađenim u dokumentima, kako bi korisnici mogli postavljati upite na suvremenom jeziku i pravopisu. Sve su to problemi koje sustav temeljen na metodama za obradu prirodnog jezika na povijesnim tekstovima treba prevladati te će se na temelju ovih zaključaka predložiti i odgovarajući model.

U posljednjih nekoliko desetljeća u području obrade prirodnoga jezika razvijeno je mnogo jezičnih resursa, jezičnih alata i ostalih kompleksnijih sustava za obradu prirodnoga jezika. Međutim, razvoj se odvijao izolirano na projektima, sveučilištima i institutima, ali i od strane pojedinaca, gdje je

nedostajalo komunikacije među članovima zajednice. Rezultat toga su resursi, alati i sustavi koji ne mogu međusobno razmjenjivati informacije.

Stoga se u najnovije vrijeme s jedne strane pokazala potreba za standardizacijom jezičnih tehnologija, ali i cijelih procesa razvoja. Interoperabilnost se javlja kao ključni pojam u ovoj fazi obrade prirodnoga jezika, čime se želi omogućiti međusobnu komunikaciju jezičnih podataka.

S druge strane, iz iskustva interoperabilnosti i suradnje uvidjelo se da automatske metode za opis prema nekoj zadanoj shemi imaju svoja ograničenja i domete te je jasno da nikad nije moguće razviti algoritam koji će potpuno točno označiti svaki tekst. Stoga je neophodan dodatni ljudski napor i vrijeme da se ti tekstovi dovedu do potpune točnosti (ili barem prihvatljive prema nekim statističkim pretpostavkama), kako bi bili prikladni za kasnije analize i obrade.

U disertaciji je u skladu s tim predložen model za automatsku odnosno poluautomatsku obradu tekstova, ispitana njegova točnost te provjereno koliko automatska ili poluautomatska obrada ubrzava proces naknadnog označavanja, odnosno dovođenja teksta do prihvatljive razine točnosti označavanja.

Kako bi se omogućila interoperabilnost jezičnih resursa, u istraživanju su korištene TEI oznake za označavanje strukture rječničkih natuknica, koje se temelje na sustavu za označavanje XML za prikaz strukture te na kodnom sustavu Unicode za prikaz znakova.

Pri obradi i označavanju rječnika treba se suočiti s dva glavna problema. Prvi je problem vezan uz strukturu natuknice, a drugi uz informaciju dostupnu u natuknici. Struktura rječničke natuknice razlikuje se od rječnika do rječnika, a najčešće i unutar jednog te istog rječnika. Kako bi shema za kodiranje mogla biti prikladna za različite strukture, jedini način je da se omogući da se bilo koji element može pojaviti bilo gdje u natuknici. Međutim, postoje i rječnici koji

imaju dosljednu strukturu, te stoga shema mora podržavati i takve slučajeve. Budući da odabrani rječnici imaju prilično dosljednu strukturu, u modelu se koristi element `entry` kojim se prikazuju strukturirane natuknice.

Drugi problem koji pri označavanju rječničkoga teksta treba razriješiti jest način predstavljanja informacija u natuknici. Većina informacija u rječnicima je implicitna ili vrlo sažeta. Stoga je potrebno odabrati treba li prikazati informacije kao u izvorniku ili implicitne informacije učiniti eksplicitnima. Budući da su istraživačima oba načina jednako potrebna i zanimljiva, shema mora podržavati i jedan i drugi način prikaza te također njihov međuodnos. Za neka je istraživanja važno zadržati izvorni prikaz natuknice, dok će za druga biti važne samo leksičke informacije što može za posljedicu imati promijenjen redoslijed pojavljivanja informacija, pri čemu je čak moguće i proširivanje, ali i izbacivanje određenih dijelova teksta. Moguća su istovremeno i oba pristupa, te je potrebno paralelno iskodirati tekst na oba načina i postaviti ih u međuodnos. Svi ti pristupi mogu se ostvariti upotrebom načela sadržanih u smjernicama TEI.

Za kodiranje tekstova natuknica starih hrvatskih rječnika najprikladniji je način zadržavanje izvornog redoslijeda pojavljivanja informacija, dok su dodatne informacije (koje se ne nalaze u izvornom tekstu) ukodirane u atributima elemenata. Na taj način razdvojit će se izvorni sadržaj, koji se pojavljuje kao sadržaj elementa, od novog sadržaja odnosno značenja, koji će se dodati u atributu elementa.

U radu su definirane oznake koje se koriste za označavanje rječničkih natuknica digitaliziranih rječnika koje se temelje na TEI oznakama, *de facto* standardu za kodiranje teksta.

Istraživanje koje se bavi primjenom poluautomatskih metoda za opis shemom provedeno je nad prvim sveskom della Bellinog talijansko-latinsko-hrvatskog rječnika iz 1785. godine tiskanog u Dubrovniku. Cilj istraživanja je automatsko označavanje jezika i strukture rječničkih natuknica.

Svezak sadrži 7 972 rječničkih natuknica od slova A do H, koje su automatski opojavničene. Svezak sveukupno sadrži 403 128 pojava, odnosno jedna rječnička natuknica sadrži 50,57 pojava. Detaljno objašnjenje što je sve određeno kao samostalna pojava može se naći u poglavlju 6.3.2.1.

Nakon što su natuknice opojavničene, iz korpusa je slučajnim odabirom izdvojena 101 natuknica za ručno označavanje. Svaka pojava unutar 101 izdvojene natuknice označena je na dvije razine: oznakom za jezik i oznakom za strukturu. Jezična razina ima tri različite oznake, dok strukturna razina ima 19 oznaka. Sveukupno je ručno označeno 8 340 oznaka u svakoj skupini, odnosno sveukupno 16 680 oznaka. To znači da u prosjeku jedna rječnička natuknica u ručno označenom uzorku sadrži 82,57 pojava, odnosno 32 pojavnice više nego prosjek natuknice u cijelom prvom svesku rječnika.

Zanimljiv je podatak da je preko polovine (53 %) pojava označeno hrvatskim jezikom, dok je talijanski jezik nešto više zastupljen nego latinski (26 % naspram 21 %), što pokazuje temeljitost autora u prikazu svih mogućih riječi sličnih značenja u hrvatskom jeziku u ovom rječniku, a time i bogatstvo hrvatskoga jezika. Jasno je da je zbog bliskosti talijanskoga i latinskoga jezika kao latinski ekvivalent talijanskoga pojma većinom uzeta samo jedna latinska riječ, dok je za hrvatski u prosjeku (prema ručnom označavanju) navedeno gotovo dvije i pol riječi po pojmu (odnosno za 1 latinsku riječ, navedeno je 2,47 hrvatskih riječi), odnosno u prosjeku su za svaki latinski prijevod navedene podjednako dvije odnosno tri ekvivalentna hrvatska prijevoda.

Za automatsko označavanje jezika nadziranom metodom strojnog učenja korišten je najsuvremeniji algoritam nadziranog strojnog učenja za označavanje sekvenci nazvan uvjetna nasumična polja (engl. *conditional random fields*, *CRF*). CRF je statistička metoda za predviđanje strukture, koja ima mogućnost predvidjeti više međusobno zavisnih varijabli. Primjenjuje se u različitim područjima poput računalnog vida, bioinformatike te, naravno, računalne obrade prirodnoga jezika.

Alat korišten za učenje i evaluaciju je *crfsuite*²¹⁶, koji implementira metodu uvjetnih nasumičnih polja. Alat implementira više različitih vrhunskih metoda strojnog učenja, a za ovo je istraživanje korišten pasivno agresivni algoritam za učenje jer je pokazao najbolje rezultate. Alat omogućava brzo učenje i označavanje podataka, jednostavan format podataka te dizajniranje proizvoljnog broja značajki za svaku jedinicu. Također je moguće ovim alatom izračunati rezultate modela evaluiranih na skupu za provjeru (preciznost, potpunost, F mjera), koje su detaljnije opisane u poglavlju 6.3.3.1.

Svaki parametar evaluiran je računajući točnost putem 10-struke unakrsne validacije (*10-fold cross-validation*). U navedenoj se metodi podskupovi za učenje grade iz skupa za učenje slučajnim uzorkovanjem.

Unakrsna validacija je metoda validacije koja omogućava procjenu kako će se rezultati statističke analize uopćiti na nezavisnom skupu podataka. Uglavnom se koristi u slučajevima kada se želi procijeniti koliko će točan biti prediktivni model nad nepoznatim podacima. Prediktivni modeli uglavnom sadrže dva skupa podataka. Jedan skup tih podataka su podaci za učenje, koji sadrži poznate podatke, dok drugi skup sadrži nepoznate podatke. Cilj unakrsne validacije je definiranje skupa podataka za provjeru modela u fazi učenja, čime

²¹⁶<http://www.chokkan.org/software/crfsuite/>

se stvara novi podskup podataka za učenje, odnosno skup za validaciju. Cilj unakrsne validacije je nastojanje da se izbjegne preučenost modela te da se vidi kako će se model uopćiti na nepoznatim podacima.

Unakrsna validacija upotrijebljena u ovom istraživanju je takozvana n -struka unakrsna validacija, gdje n iznosi 10. Kod ove vrste unakrsne validacije izvorni uzorak je nasumično podijeljen na n (odnosno 10) podskupova jednakih veličina. Jedan se podskup koristi kao skup za validaciju za testiranje modela, dok se ostalih $n-1$ (odnosno 9) podskupova koristi kao skupovi za učenje. Zatim se proces unakrsne validacije ponavlja n puta, gdje se svaki od n podskupova koristi točno jednom kao podaci za validaciju. Dobiveni su rezultati uprosječeni kako bi se dobila jedna procjena.

Problemu označavanja jezika pojava i problemu označavanja strukture pristupilo se jednako. Prvo su definirani potencijalno zanimljivi skupovi značajki koji bi mogli dati bolje rezultate od samih podataka. Zatim je izvršeno mjerenje uspješnosti skupova značajki. Konačno, objedinjeni su svi korisni skupovi značajki kako bi se postigao najbolji mogući rezultat s definiranim značajkama.

Istraživanje je provedeno u tri faze. Prva faza sastoji se od provjere najočitiije značajke, odnosno utječeli način pisanja natuknice na rezultat: izvorni oblik pojava i njezin oblik pretvoren u kurent (pisano sve malim slovima). Očekuje se da će jedan od ta dva načina pisanja natuknice ostvariti bolji rezultat. Stoga se značajka s ostvarenim boljim rezultatom upotrebljava u sljedećim fazama istraživanja kao osnovna značajka. Kod označavanja jezika pojava pretvorene u kurent postižu bolje rezultate od izvornog oblika pojava (0,94143 naspram 0,93224). Međutim, kod označavanja strukture rezultat je suprotan, odnosno izvorni oblik pojava postiže bolje rezultate od pojava pretvorenih u kurent (0,85993 naspram 0,85993).

U drugoj fazi istraživanja provjerava se utjecaj dodatnih značajki na rezultatestrojnog učenja. Kao osnovna značajka koristi se ona iz prve faze. Kod jezičnih oznaka kao dodatna značajka mjerena je logička varijabla kojom je zabilježen podatak je li originalna pojava pisana kurentom ili nije. Zatim je mjerena frekvencija pojavljivanja određenog troslova. Nadalje je provjeren utjecaj prethodnih i sljedećih pojava, gdje se N kreće od 1 do 3. Konačna značajka mjerena u eksperimentu je logička varijabla kojom je zabilježen podatak je li prethodna ili sljedeća pojava pisana kurentom ili nije.

Kod strukturnih oznaka kao dodatna značajka mjerena je logička varijabla kojom je zabilježen podatak je li originalna pojava pisana kurentom ili nije. Zatim je provjeren utjecaj prethodnih i sljedećih pojava, gdje se N kreće od 1 do 4. Nadalje, značajka upotrijebljena u eksperimentu je logička varijabla kojom je zabilježen podatak je li prethodna ili sljedeća pojava pisana kurentom ili nije. Budući da u ovom skupu podataka svaka pojava sadrži podatak o jeziku, provjeren je utjecaj te značajke na rezultat. Konačno, provjeren je utjecaj sufiksa na rezultat, gdje su uzeta posljednja četiri znaka pojave.

U trećoj fazi istraživanja, sveznačajke koje pokazuju bolje rezultate naspram rezultata postignutog osnovnom značajkom objedinjenesu u jedan eksperiment. Time se postiže najbolji rezultat s definiranim značajkama.

Konačno, sveznačajke koje pokazuju bolje rezultate naspram rezultata postignutog osnovnom značajkom objedinjene su u jednu cjelinu. Ipak, i dalje postoji prostor za unaprjeđenje u potrazi za najpovoljnijim skupom značajki, i to definiranjem novih značajki. Kod označavanja jezika postignuta je točnost od 0,98413, a kod označavanja strukture točnost iznosi 0,96371.

Sljedećim je eksperimentom odgovoreno na pitanje je li ispravljanje automatskih oznaka za jezik i strukturu vremenski brže naspram ručnog

označavanja, te ako je odgovor potvrđan, koliko je brže. Rezultat eksperimenta potvrđuje da je ispravljanje 4,64 puta brže nego unošenje oznaka. Dodatna vrijednost ovog eksperimenta jest 4 180 pojava koje su naknadno ručno pregledane. Ispravljanjem materijala možemo povećavati skup za učenje čime bi se dobili točniji modeli, a iz toga proizlazi da bi se time još više smanjilo vrijeme ispravljanja.

Završnim eksperimentom na skupu za provjeru detaljnije je analizirana uspješnost klasifikatora prikazom tablice zabuna pomoću koje se može vidjeti koje su to kategorije koje algoritam često pogrešno označava. Detaljna analiza nalazi se u poglavlju 6.3.7. Točnost algoritma za jezične oznake je 0,97308, a za strukturne oznake 0,954801, što su vrlo zadovoljavajući rezultati.

Analizom krivulje učenja ustanovljeno je da prepoznavanje jezika ima bolje rezultate nego prepoznavanje strukture (98,59 % naspram 95,92 %). Budući da jezični skup ima samo tri oznake, a strukturni 19, rezultati su očekivani. Krivulja jezičnih oznaka brže raste te je već nakon 20 % podataka (odnosno oko 79 000 pojava) postignuta točnost od gotovo 96 %. Strukturne oznake postižu slabije rezultate te je učenje sporije. Potrebno je otprilike 40 % podataka (odnosno oko 158 000 pojava) kako bi se postigla točnost veća od 94 %. Oba rezultata se smatraju zadovoljavajućima. Budući da krivulje za obje razine još značajno rastu, moguće je popraviti izlazne rezultate trenutnog algoritma, tako da se uključe dodatni i novi podaci.

Zahvaljujući formaliziranom prikazu rječničkih natuknica putem TEI oznaka, moguće je mapirati elemente na LMF (*Lexical Markup Framework*). Označavanje resursa dodatnim standardom proširuje potencijalne zainteresirane pojedince i projekte, koji bi se mogli nadovezati na resurs.

Osim uvođenja novog standarda, koristan bi bio sustav pretraživanja povijesnih dokumenata koji bi omogućio uparivanje upita na suvremenom jeziku s povijesnim oblicima pronađenim u dokumentima. Na taj bi način korisnici mogli postavljati upite na suvremenom jeziku i pravopisu.

Koncepte unutar rječnika moguće je prikazati putem RDF-a (*Resource Description Network*). Postoje pokušaji mapiranja LMF-a i RDF-a, koje valja istražiti te vidjeti postoji li mogućnost automatskog mapiranja na RDF.

Konačno, moguće je korištenje mnoštva (*crowdsourcing*), kako za provjeru točnosti teksta za vrijeme digitalizacije tako i za povezivanje s vanjskim resursima i konceptima.

Znanstveni doprinos istraživanja je višestruk. S jedne strane, formaliziran je problem digitalizacije i obrade povijesnih rječničkih tekstova. S druge strane, razvijen je model višerazinskog prikaza tekstova starijih hrvatskih rječnika, koji omogućava daljnja istraživanja ne samo u informacijskim znanostima, točnije računalnoj leksikografiji kao dijelu informacijskih znanosti, nego i u ostalim lingvističkim i filološkim, ali i kulturološkim te povijesnim područjima i granama znanosti. Konačno, primijenjene su metode za automatsku i poluautomatsku obradu povijesnih tekstova koji sadrže različita dijakronijska stanja jezika.

Ciljevi istraživanja su postignuti. Omogućen je prikaz znanja u starijim rječničkim tekstovima. Također je omogućena interoperabilnost s drugim jezičnim resursima, alatima i sustavima za obradu prirodnoga jezika. Konačno, ubrzan je i pojednostavljen proces obrade starijih rječničkih tekstova.

8. Literatura

1. **Arhinet, arhivski informacijski sustav.** Hrvatski državni Arhiv. Zagreb. <http://arhinet.arhiv.hr/> (26. 8. 2014.).
2. **Bacchiani, M. et al.** MAP adaptation of stochastic grammars. // Computer Speech and Language. 20, 1(2006), str. 41-68.
3. **Gabrić-Bagarić, Darija.** Četiri ishodišta hrvatskoga standardnoga jezika. Rijeka. Odsjek za kroatistiku Filozofskoga fakulteta Sveučilišta u Rijeci. Fluminensia 22 (2010), br. 1. str 149-162.
4. **Bago, Petra.** „Označavanje strukture rječnika teksta prvog sveska drugog izdanja della Bellinog talijansko-latinsko-hrvatskog rječnika iz 1785.“Diplomski rad, Faculty of Humanities and Social Sciences, University of Zagreb, 2008.
5. **Bago, Petra; Kolarec, Martina.** „Računalna priprema, obrada i prikaz sedmojezičnog rječnika Petera Lodereckera povodom 400. godišnjice izdavanja.“ Studentski znanstveni rad, Rektorova nagrada, Filozofski fakultet, Sveučilište u Zagrebu, 2006.
6. **Bakhilid** // Wikipedija, Slobodna enciklopedija. <http://sh.wikipedia.org/wiki/Bakhilid> (2. 9. 2014.).
7. **Bartol Kašić** // Wikipedija, Slobodna enciklopedija. http://hr.wikipedia.org/wiki/Bartol_Kašić (31. 8. 2014.).
8. **Berners-Lee, Tim.** „Linked Data - Design Issues,“ July 27, 2006. <http://www.w3.org/DesignIssues/LinkedData.html>.
9. **Betta Code** // Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Beta_Code (2.9. 2014.).

10. **Bizer, Christian; Lehmann, Jens; Kobilarov, Georgi; Auer, Sören; Becker, Christian; Cyganiak, Richard; Hellmann, Sebastian.**„DBpedia - A Crystallization Point for the Web of Data.“ Web Semant. 7 (2009): 154-165.

11. **Boras, Damir.**„Croatian dictionary heritage - collection of old Croatian dictionaries in digital form.“Conference on Journalism and Information Technology and UNESCO project on the word heritage Colloquium on the project „Croatian dictionary heritage and dictionary knowledge presentation“Inter-University Centre, Dubrovnik, May 26-28, 2004 (pozvano predavanje, neobjavljeni rad) <http://bib.irb.hr/prikazi-rad?&rad=182779>.

12. **Boras, Damir.**„Teorija i pravila segmentacije teksta na hrvatskom jeziku.“, Doktorska disertacija, Zagreb, Sveučilište u Zagrebu Filozofski fakultet, 1998. str 284.

13. **Boras, Damir (ur).** Portal hrvatske rječničke baštine – Croatian Old Dictionary Portal, Zagreb, 2005, <http://cro dip.ffzg.hr/>.

14. **Boras, Damir; Ljubešić, Nikola; Mikelić, Nives.** Croatian Old Dictionary Portal (CRODIP) // LIDA 2005 Proceedings: Libraries in the digital age 2005, 30.5.-3.6.2005, Dubrovnik-Mljet, Hrvatska, Faculty of Philosophy, University J. J. Strossmayer in Osijek, Croatia: Rutgers University, New Brunswick, NJ, USA, 2005., 127-137.

15. **Boras, Damir; Mikelić, Nives.**Rječnik Fausta Vrančića – temelj hrvatske rječničke baštine (računalna obradba)// Modeli znanja i obrada prirodnoga jezika / Tuđman, Miroslav (ur.). Zagreb : Zavod za informacijske studije, Filozofski fakultet, 2003. Str. 237-272.

16. **Boras, Damir** (voditelj). Hrvatska rječnička baština i prikaz rječničkoga znanja (0130464); Znanstveni projekt (MZOS RH), Sveučilište u Zagrebu Filozofski fakultet, Zagreb, 2003-2006.

http://zprojekti.mzos.hr/public/c2prikaz_det.asp?cid=1&psid=25&ID=451. (28. 8. 2014.). Bibliografija na: http://bib.irb.hr/lista-radova?sif_proj=0130464

17. **Boras, Damir** (voditelj). Hrvatska rječnička baština i hrvatski europski identitet (130-1301679-1380); Znanstveni projekt (MZOS RH); Sveučilište u Zagrebu Filozofski fakultet, Zagreb, 2007-2013. http://zprojekti.mzos.hr/public/c2prikaz_det.asp?cid=1&psid=25&ID=451. (28. 8. 2014.). Bibliografija na: http://bib.irb.hr/lista-radova?sif_proj=130-1301679-1380&period=2007

18. **Boras, Damir; Tadić, Marko**. Dva značajna projekta izgradnje računalnih resursa za hrvatski jezik // Thesaurus Archigymnasii, Zbornik radova u prigodi 400. godišnjice Klasične gimnazije u Zagrebu / Koprek, Ivan (ur.). Zagreb : Klasična gimnazija u Zagrebu, 2007. Str. 546-559.

19. **Bray, Tim; Paoli, Jean; Sperberg-McQueen, C. M.; Maler, Eve ; Yergeau, François ; Cowan, John.**, „Extensible Markup Language (XML).“ World Wide Web Journal 2, no. 4 (1997): 27-66.

20. **Brod**. 2013. *Hrvatski Jezični Portal*. Novi Liber. http://hjp.novi-liber.hr/index.php?show=search_by_id&id=f15kURY%3D&keyword=brod. (3.2.2013.).

21. **Chambers's Twentieth Century Dictionary of the English Language** (part 1 of 4: A-D), London and Edinburgh, 1908. <http://www.gutenberg.org/files/37683/37683-h/37683-h.htm> (30.8.2014.).

22. **Confusion matrix** // Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Confusion_matrix (8.9.2014.).

23. **Cross-validation (statistics)** // Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Cross-validation_%28statistics%29 (7.9.2014.).

24. **Čubranović, Andrija.** 2014. *Proleksis Enciklopedija*. Vujić, Antun (ur.). Leksikografski zavod Miroslav Krleža. <http://proleksis.lzmk.hr/16578/>. (1.9.2014.).
25. **Čučić, V.** Prvi tiskari u Dubrovniku s popisom tiskane grade. // *Vjesnik bibliotekara Hrvatske*. 48, 3/4(2005), str. 108-158. (str. 119).
26. **DBpedia.** <http://dbpedia.org/About>. (5.9.2014.),
27. **della Bella, Ardelio.** *Dizionario Italiano-latino-illirico: a Cui Si Premettono Alcune Brevi Istruzioni Gramaticali, Necessarie Per Apprendere La Lingua e L'ortografia Illirica*. 2. izdanje. Dubrovnik, 1785.
28. **Bunz, Carl-Martin.** *Der Thesaurus indogermanischer Text- und Sprachmaterialien (TITUS) – ein Pionierprojekt der EDV in der Historisch-Vergleichenden Sprachwissenschaft*. U: *Sprache und Datenverarbeitung* 1, 1998, 10-30.
29. **Daume III, H.** Frustratingly easy domain adaptation. // *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Prag : Association for Computational Linguistics, 2007. Str. 256-263. <http://www.umiacs.umd.edu/~hal/docs/daume07easyadapt.pdf>. (4.9.2014.).
30. **Digitalizirati.** 2013. *Hrvatski Jezični Portal*. Novi Liber. http://hjp.novi-liber.hr/index.php?show=search_by_id&id=f1dmXxQ%3D&keyword=digitalizirati. (3.2.2013.).
31. **Digitize.** 2010. *Oxford Dictionaries*. Oxford University Press. <http://oxforddictionaries.com/definition/english/digitize?q=digitise>. (3. 2. 2013.).
32. **Divković, Mirko.** *Latinsko-hrvatski Rječnik Za Škole*. 2. izdanje. Zagreb: Nakladom Kr. hrvatsko-slavonsko-dalmatinske zemaljske vlade, 1900. (Reprint: 3. izdanje, Naprijed. Zagreb, 1980.)

33. **Dominiko Zlatarić** // Wikipedija, Slobodna enciklopedija. http://hr.wikipedia.org/wiki/Dominko_Zlatarić (31. 8. 2014.).
34. **Enciklopedija Leksikografskog Zavoda Miroslav Krleža**, Mrežno izdanje. <http://www.enciklopedija.hr/> (5.9.2014.).
35. **Encyclopedia of Machine Learning**. Sammut, Claude; Webb, Geoffrey I. (ur.). New York : Springer, 2011.
36. **Engleski Jezik**. 2013. *Proleksis Enciklopedija*. Vujić, Antun (ur.). Leksikografski zavod Miroslav Krleža. <http://proleksis.lzmk.hr/55584/>. (4.2.2013.).
37. **eXtensible Markup Language (XML)**. <http://www.w3.org/XML/>. (10.7.2014.).
38. **Fostering Language Resources Network (FLaReNet)**. <https://www.flarenet.eu/>. (5.9.2014.).
39. **Francuski Jezik**. 2013. *Proleksis Enciklopedija*. Vujić, Antun (ur.). Leksikografski zavod Miroslav Krleža. <http://proleksis.lzmk.hr/667/>. (4.2.2013.).
40. **Gildea, D.** Corpus variation and parser performance. // Proceedings of EMNLP / uredili L. Lee i D. Harman. Pittsburgh, 2001. Str. 167-202. <http://www.cs.rochester.edu/~gildea/gildea-emnlp01.pdf>. (4. 9. 2014.).
41. **Gortan, Veljko**. (O latinsko-hrvatskom rječniku Mirka Divkovića). Predgovor trećem reprint izdanju latinsko-hrvatskog Rječnika za škole Mirka Divkovića iz 1900. Reprint: Naprijed, Zagreb, 1980. str. IV-V.
42. **Gostl, Igor**. *Dragutin Antun Parčić*. Mala Knjižnica Matice Hrvatske, kolo 8, knj. 47. Zagreb: Matica hrvatska, 1998.
43. **Hilbert, Martin; López, Priscila**. 2011. „The World’s Technological Capacity to Store, Communicate, and Compute

Information.“*Science* 332 (6025) (1.4.2013.): 60-65.
doi:10.1126/science.1200970.

44. **Homonimija.** 2013. *Hrvatski Jezični Portal*. Novi Liber.
[http://hjp.novi-liber.hr/index.php?show=search_by_id&id=fVxgXxY%3D](http://hjp.novi-liber.hr/index.php?show=search_by_id&id=fVxgXxY%3D&keyword=homonimija)
&keyword=homonimija. (3.2.2013.).

45. **Horvat, Vladimir.** Jakov Mikalja, Isusovac–leksikograf. Zagreb,
Filozofsko teološki Institut Družbe Isusove. Obnovljeni život (56) 4 (2001) str.
495-502.

46. **Hrvatski državni arhiv,**
<http://www.arhiv.hr/arhiv2/Onama/index.htm> (3. 9. 2014.).

47. **Hrvatski Jezični Portal.** <http://hjp.novi-liber.hr/>. (5.9.2014.).

48. **Hrvatski Jezik.** 2013. *Proleksis Enciklopedija*. Vujić, Antun (ur.).
Leksikografski zavod Miroslav Krleža. <http://proleksis.lzmk.hr/27217/>.
(4.2.2013.).

49. **Ide, Nancy; Pustejovsky, James.**„What Does Interoperability
Mean, Anyway? Toward an Operational Definition of Interoperability,“ 2010.
<http://www.cs.vassar.edu/~ide/papers/ICGL10.pdf>.

50. **Ide, Nancy; Pustejovsky, James; Calzolari, Nicoletta; Soria,
Claudia.** The SILT and FLReNet International Collaboration for
Interoperability. Proceedings of the Third Linguistic Annotation Workshop,
ACL-IJCNLP 2009, str. 178-181.

51. **Ignjat Đurđević** // Wikipedija, Slobodna enciklopedija.
http://hr.wikipedia.org/wiki/Ignjat_Đurđević (31.08.2014.).

52. **Jambrešić, Andrija.**Lexicon Latinum Interpretatione Illyrica,
Germanica, Et Hungarica Locuples, ... Digestum, Ab Andrea Jambressich.
Zagrabiae: Typis Academicis Societatis Jesu, per Adalbertum Wilh. Wesseli,
1742.

53. **Korade, M.; Aleksic, M.; Matoš, J.** Isusovci i hrvatska kultura. Zagreb : Hrvatski povijesni institut u Beču, 1993. Str. 49-52, 269-273.

54. **Lafferty, J., McCallum, A., Pereira, F.** (2001). "Conditional random fields: Probabilistic models for segmenting and labeling sequence data". Proc. 18th International Conf. on Machine Learning. Morgan Kaufmann. pp. 282-289.

55. **Lak (1).** 2013. *Hrvatski Jezični Portal*. Novi Liber. http://hjp.noviliber.hr/index.php?show=search_by_id&id=elZiXxA%3D&keyword=lak. (3.2.2013.).

56. **Lak (2).** 2013. *Hrvatski Jezični Portal*. Novi Liber. http://hjp.noviliber.hr/index.php?show=search_by_id&id=elZiXxE%3D&keyword=lak. (3.2.2013.).

57. **Lassila, Ora; Swick, Ralph R.M; World Wide Web Consortium.** Resource Description Framework (RDF) Model and Syntax Specification, 1998.

58. **Lavrenko, Victor.** Text classification 3: Passive Aggressive Algorithm. Objavljeno 20.1.2014. http://www.youtube.com/watch?v=TJU8NfDdqNQ&index=3&list=PLBv09BD7ez_4XyTO5MnDLV9N-s6kgXQy7 (7.9.2014.).

59. **Lexical Markup Framework** // Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Lexical_Markup_Framework (5.9.2014.).

60. „Linked Data - Connect Distributed Data Across the Web.“ <http://linkeddata.org/>. (19.2.2013.).

61. **Lisac, Josip.** Držićev jezik i leksikografija njegova doba // Čakavska rič. 35, 1(2007) str. 9-14.

62. **Loderecker, Petr.** Dictionarium Septem Diversarum Linguarum, Videlicet Latine, Italice, Dalmatice, Bohemice, Polonice, Germanice & Ungarice Una Cum Cuiuslibet Linguae Registro Sive Repertorio Vernaculo in

Quo Candidus Lector Sui Idiomatis Vocabulum Facile Invenire Poterit. Cum gratia & privilegio imperiali. Prag: apud Nicolaum Straus, e typographeo Ottmariano, impensis ac sumptibus authoris, 1605.

63. **Ljubešić, Nikola.** „Pronalaženje događaja u višestrukim izvorima informacija.“ Doktorska disertacija. Filozofski fakultet, Sveučilište u Zagrebu, 2009.

64. **Mažuranić, Ivan i Jakov Užarević.** Deutsch-ilirisches Wörterbuch. Njemačko-ilirski slovar, Verlag und Druck der k. priv. Ilirischen National-Buchdruckerei von Dr. Ljudewit Gaj. Zagreb, 1842.

65. **McClosky, D.; Charniak, E.; Johnson, M.** Reranking and self-training for parser adaptation. // Proceedings of COLING-ACL. Sydney : Association for Computational Linguistics, 2006. Str. 337-344. <http://nlp.stanford.edu/~mcclosky/papers/dmcc-acl-2006-slides.pdf>. (4. 9. 2014.).

66. **Mikalja, Jakov.** Blago Jezika Slovinskoga Illi Slovník u Komu Izgorarajuse Rjeci Slovinske Latinski i Diacki = Thesaurus Linguae Illyricae Sive Dictionarium Illyricum in Quo Verba Illyrica Italice et Latine Redduntur. Loreto/Ancona: et sumptibus Sacrae congregationis de propaganda fide impressum ; apud Paulum et Io. Baptistam Seraphinum, 1649.

67. **Mikalja, Jakov.** *Hrvatska Enciklopedija, Mrežno Izdanje*. Ravlić, Slaven (ur.). Zagreb : Leksikografski zavod Miroslav Krleža. <http://www.enciklopedija.hr/Natuknica.aspx?ID=40703> (29.8.2014.).

68. **Mitchell, Tom M.** Machine learning. New York : McGraw Hill, 1997.

69. **Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet.** Foundations of Machine Learning. MIT Press, 2012.

70. **Monumenta antiquissima**, HR-HDA-876-1, 999, 15. srpnja, U Zadru http://arhinet.arhiv.hr/_DigitalniArhiv/Monumenta/HR-HDA-876-1.htm (3. 9. 2014.).

71. **Nacionalna i sveučilišna knjižnica u Zagrebu**, <http://www.nsk.hr/> (3. 9. 2014.).

72. **Nikola Dimitrović** // Wikipedija, Slobodna enciklopedija. http://hr.wikipedia.org/wiki/Nikola_Dimitrović (29. 8. 2014.).

73. **Njemački Jezik**. 2013. *Proleksis Enciklopedija*. Vujić, Antun (ur.). Leksikografski zavod Miroslav Krleža. <http://proleksis.lzmk.hr/59159/>. (4.2.2013.).

74. **Okazaki, Naoaki**. CRFSuite : a fast implementation of Conditional Random Fields (CRFs). 2007. <http://www.chokkan.org/software/crfsuite/> (29. 7.2014.).

75. **Parčić, Dragutin Antun**. Rječnik Hrvatsko-talijanski. 3. izdanje. Zadar: Tipografia editrice Narodni list, 1901.

76. **Parčić, Dragutin Antun**. *Hrvatska Enciklopedija, Mrežno Izdanje*. Ravlić, Slaven (ur.). Zagreb : Leksikografski zavod Miroslav Krleža. <http://www.enciklopedija.hr/Natuknica.aspx?ID=46672> (29. 8. 2014.).

77. **Perseus Project** // Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Perseus_Project (2.9.2014.).

78. **Petar Hektorović** // Wikipedija, Slobodna enciklopedija. http://hr.wikipedia.org/wiki/Petar_Hektorović (29. 8. 2014.).

79. **Piotrowski, Michael**. Natural Language Processing for Historical Texts. Edited by Graeme Hirst. Synthesis Lectures on Human Language Technologies 17. Morgan & Claypool Publishers, 2012. <http://www.morganclaypool.com/doi/abs/10.2200/S00436ED1V01Y201207HLT017>.

80. **Polisemija.** 2013. *Hrvatski Jezični Portal.* Novi Liber. http://hjp.noviliber.hr/index.php?show=search_by_id&id=eVtgWRA%3D&keyword=polisemija. (2.3.2014.).

81. **Povijest.** 2013. *Proleksis Enciklopedija.* Vujić, Antun (ur.). Leksikografski zavod Miroslav Krleža. <http://proleksis.lzmk.hr/42422/>. (4. 2. 2013.).

82. **Perseus Digital Library,** Tufts University. <http://www.perseus.tufts.edu/hopper/> (2.9.2014.).

83. **Precision and recall** // Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Precision_and_recall (7.9.2014.).

84. **Prirodni jezik.** 2013. *Hrvatski Jezični Portal.* Novi Liber. http://hjp.noviliber.hr/index.php?show=search_by_id&id=fVZvWhc%3D&keyword=jezik. (3.2.2013.).

85. **Project Gutenberg.** <http://www.gutenberg.org/> (25.8.2014.).

86. **Project Gutenberg** // Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Project_Gutenberg (25. 8. 2014.).

87. **Projekt digitalizirane baštine Nacionalne i sveučilišne biblioteke,** Zagreb, 2013. <http://db.nsk.hr/> (26.8.2014.).

88. **Projekt eLektire,** <http://lektire.skole.hr/> (3.9.2014.).

89. **Projekt TITUS,** tekst Odiseje. <http://titus.uni-frankfurt.de/texte/etcs/grie/homer/odyssee/odyss.htm> (3.9.2014.).

90. **Proleksis Enciklopedija.**” <http://proleksis.lzmk.hr/>. (5. 9. 2014.).

91. **Radna grupa za digitalizaciju arhivske, knjižnične i muzejske građe Ministarstva kulture Republike Hrvatske.** 2006. „Prijedlog Nacionalnog programa digitalizacije arhivske, knjižnične i muzejske građe“.

Ministarstvo kulture Republike Hrvatske.
<http://daz.hr/bastina/NacionalniProgramDigitalizacije.pdf>. (5.2.2013.).

92. **reStructuredText** // Wikipedia, The Free Encyclopedia.
<http://en.wikipedia.org/wiki/ReStructuredText> (30. 8. 2014.).

93. **RDF** - Semantic Web Standards.” <http://www.w3.org/RDF/>.
(5. 9. 2014.).

94. **Romary, Laurent; Wegstein, Werner.** „Consistent Modeling of Heterogeneous Lexical Structures“, Journal of the Text Encoding Initiative [Online], Issue 3 | November 2012, Online since 15 October 2012, connection on 15 April 2014. URL : <http://jtei.revues.org/540> ; DOI : 10.4000/jtei.540

95. **Seiter-Šverko, Dunja, Križaj, Lana.** „Digitalizacija kulturne baštine u Republici Hrvatskoj: od trenutne situacije prema nacionalnoj strategiji“ presented at the 2. NAPLE konferencija: Uloga narodnih knjižnica u nacionalnim politikama izgradnje društva znanja (Digitalizacija kulturne baštine i provedba multikulturalizma u knjižnicama), 2005. August 5, Supetar, Hrvatska. <http://www.min-kulture.hr/naple/hrv/index.htm>.

96. **Sekine, S.** The domain dependence of parsing. // Proceedings od Applied Natural Language Processing (ANLP). 1997. Str. 96-102.
<http://ww2.cs.mu.oz.au/acl/A/A97/A97-1015.pdf>. (4.9.2014.).

97. **Sérasset, Gilles** (2012). DBnary: Wiktionary as a LMF based Multilingual RDF network. *In Language Resources and Evaluation Conference, LREC 2012*, Istanbul.

98. **Sironić-Bonefačić, Nives.** „Analiza talijanske i hrvatske leksičke građe u rječniku Ardelija Della Belle“ Doktorska disertacija. Filozofski fakultet, Sveučilište u Zagrebu, 1986.

99. **Sironić-Bonefačić, Nives.**„DELLA BELLA, Ardelio.“ Macan, Trpimir (ur.). Zagreb: Leksikografski zavod Miroslav Krleža, 1993. <http://hbl.lzmk.hr/clanak.aspx?id=4507>.

100. **Sustainable Interoperability for Language Technology (SILT).** <http://www.anc.org/LAPPS/SILT/index.html>. (5.9.2014.).

101. **Sutton, Charles; McCallum, Andrew.** An Introduction to Conditional Random Fields. //Foundations and Trends in Machine Learning. 4, 4(2011), str. 267-373. <http://homepages.inf.ed.ac.uk/csutton/publications/crftut-fnt.pdf> (7.9.2014.).

102. **Stulli, Joakim.** Rječosložje Iliričko-italijansko-latinsko. Dubrovnik, 1806.

103. **The Perseus Digital Library**, <http://www.perseus.tufts.edu/hopper/> (2. 9. 2014.).

104. **Thesaurus Indogermanischer Text- und Sprachmaterialien** // Wikipedia, Die freie Enzyklopädie. http://de.wikipedia.org/wiki/Thesaurus_Indogermanischer_Text-_und_Sprachmaterialien (31.8.2014.).

105. **Johan Wolfgang Goethe Universität, Frankfurt am Main.** Thesaurus Indogermanischer Text- und Sprachmaterialien <http://titus.uni-frankfurt.de/indexe.htm>. (31.8.2014.).

106. **Tufts University**, privatno istraživačko sveučilište, Medford/Somerville, blizu Boston, Massachusetts, SAD. <http://www.tufts.edu/> (2. 9. 2014.).

107. **Titus (project)** // Wikipedia, The Free Encyclopedia. [http://en.wikipedia.org/wiki/TITUS \(project\)](http://en.wikipedia.org/wiki/TITUS_(project)) (2.9.2014.).

108. **Turk, Marija.**„HRVATSKI PRIJEVODNI EKVIVALENTI STRUČNOG NAZIVLJA U PARČIĆEVIM RJEČNICIMA.“*FLUMINENSIA: Časopis Za Filološka Istraživanja* 18, no. 1 (Studen 2006): 41-51.

109. **Turk, Marija.**, „Stranojezični Elementi U Rječnicima Dragutina Parčića.“ *Filologija*, no. 59 (Ožujak 2013): 195-205.

110. **Vajs, Nada.**, „FITONIMIJA U PARČIĆEVU RJEČNIKU HRVATSKO-TALIJANSKOME IZ 1901.“ *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje* 29, no. 1 (2003): 339-53.

111. **Vince, Zlatko.**, „Leksikograf Dragutin Parčić U Svojem Vremenu.“ In *Rječnik Hrvatsko-Talijanski*, by Dragutin Antun Parčić, 1241-76. Zagreb: Artresor studio, 1995.

112. **Vrančić, Faust.** *Dictionarium Quinque Nobilissimarum Europae Linguarum: Latinae, Italicae, Germanicae, Dalmatiae [!] et Ungaricae; Vocabula Dalmatica Quae Ungari Sibi Usurparunt; Institutio Christiana; Symbolum Apostolorum. Cum privilegiis.* Venecija: apud Nicolaum Morettum, 1595.

113. **Vrančić, Faust.** *Hrvatska Enciklopedija, Mrežno Izdanje.* Ravlić, Slaven (ur.). Zagreb : Leksikografski zavod Miroslav Krleža. <http://www.enciklopedija.hr/Natuknica.aspx?ID=65386> (29.8.2014.).

114. **Word-sense disambiguation** // Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Word-sense_disambiguation (4. 9. 2014.).

115. **Žubrinić, Krunoslav.**, „Automatsko stvaranje konceptualne mape iz nestrukturiranoga teksta na hrvatskome jeziku“. Doktorska disertacija, Fakultet elektrotehnike i računarstva, Sveučilište u Zagrebu, 2014.

9. Sažetak

Cilj istraživanja je razvoj modela koji omogućavavišerazinski prikaz znanja u tekstovima starijih hrvatskih rječnika. Dodatno je omogućena interoperabilnost s drugim jezičnim resursima, alatima i sustavima za obradu prirodnoga jezika. Model prikaza znanja u rječnicima proveden je nad sedam odabranih rječnika tiskanih u rasponu od preko 300 godina (1595. – 1901.). Koristeći *de facto* standard (*Text Encoding Initiative, TEI*), omogućena je interoperabilnost resursa.

Konačno, primijenjene su metode za automatsku i poluautomatsku obradu digitaliziranih povijesnih tekstova čime se ubrzava i pojednostavljuje proces obrade starijih rječničkih tekstova. Korišten je najsuvremeniji algoritam nadziranog strojnog učenja za označavanje sekvenci nazvan uvjetna nasumična polja (engl. *conditional random fields, CRF*). Istraživanje je provedeno na jednom rječniku s najsloženijom strukturom rječničke natuknice. Skup podataka sadrži 7 972 rječničke natuknice (403 128 pojava). Skup za učenje sastoji se od 101 nasumično odabrane rječničke natuknice (8 340 pojava). Svaka pojava označena je na dvije razine: oznakom za jezik i oznakom za strukturu. Jezična razina ima tri različite oznake, dok strukturna razina ima 19 oznaka. Kod označavanja jezika postignuta je točnost od 0,98413, a kod označavanja strukture točnost iznosi 0,96371. Dodatnim eksperimentom potvrđeno je da je ispravljanje automatskih oznaka 4,46 puta brže od ručnog označavanja.

Ključne riječi: povijesni rječnici, označavanje jezika, označavanje strukture, nadzirano strojno učenje, Text Encoding Initiative, uvjetna nasumična polja

10. Summary

The aim of this research is to develop a multilevel presentation model of old Croatian dictionary texts. We enabled interoperability with other language resources, tools and systems for natural language processing. The presentation model is conducted on seven selected dictionaries printed from 1595 to 1901. By using a *de facto* standard (Text Encoding Initiative), we enabled interoperability of the resource.

Finally, we applied automatic and semiautomatic natural language processing methods for digitized historical texts thereby speeding up and simplifying the process of processing old dictionary texts. We use the state-of-the-art supervised machine learning algorithm for sequence annotation called conditional random fields (CRF). This phase of the research is conducted on one dictionary with the most complex structure of the dictionary entries. The dataset contains 7,972 dictionary entries (403,128 tokens). The training set consists of 101 randomly selected dictionary entries (8,340 tokens). We labeled each token on two levels: a language annotation and a structural annotation. The language level has three labels, while the structural annotation has 19 labels. We reach accuracy of 98.413 % for language annotation and 96.371 % for structural annotation. Additional experiment confirmed that only correcting generated labels is roughly 4.46 times faster than full manual annotation.

Keywords: historical dictionaries, language annotation, structural annotation, supervised machine learning, Text Encoding Initiative, conditional random fields

PRILOZI

11. Popis natuknica izdvojenih za ručno označavanje

11.1. Izdvojene natuknice u obliku priređenom za obradu s vidljivim oznakama

Ukupno je za ručno označavanje izdvojena 101 natuknica iz rječnika te ih ovdje navodimo u cijelosti.

Abbarbicare , per metter radici . V. Barba-|re .
Radicare .
Abbondantiffimamente . avv. <i>Abundantiffime
.</i>◊Priobilato . Priobilno . Pribogato .
Acchetatiffimo . <i>Sedatiffimus , a . m.</i> Priuta-
|şcen , na , no . Primiran , irna , irno
.◊Najutaşcenji , a , e . Najsmirenji , a , e .
Acetire , ed inacetire , farfi aceto . <i>Acefco
,◊fcis .</i> Uşoftitife , oftımfe , tiòsàmfe .
Uşce-|chife , uşcişcèmfe uşcegaòsàmfe .
Şcighnuti-|fe , ghgnivàmfe , şcighnuòsàmfe .
Uşgljuti-|tife , gljutımfe , gljutiòsàmfe .
Proşuknuti ,◊gnivam , proşuknuòfam . Pochi na
ozat .<pb>Pochi na şcizzu , fi dice del vino .
Ace-|tire del pane . Uşkvafitife , fivàmfe ,

uškva-|fiòsàmfe . Prikiifnuti , fgnivam ,
 prikiifnuò-|fam . Acetirfi del latte . Uşvifctàtfe
 , fctu-|jèmfe , uşvifctaòsàmfe . Proşuknuti ,
 gni-|vam , proşuknuòfam . v. g. Mljekòje pro-
 |şuknulo . Uşkisjelitfe , livàmfe ,
 uşkifelilòsàm-|fe . Far acetire . <i>In acetum
 convertere .</i> Ci-|niti uşoftiti . Ciniti ,
 dàfe uşofti . Uşkvafiti◊kruh ; far acetir il pane
 . Ukisjeliti , far◊acetir il latte . L' acetire ,
 il farfi aceto .◊<i>In acetum mutatio .</i>
 Uşoftegne , egna . n.◊Il farfi aceto del pane .
 Uşkvafegne , gna . n.

Adiratamente . V. Iratamente .

Affreddare . V. Raffreddare .

Affrica , ed Africa , una delle quattro parti
 del◊Mondo . <i>Africa , cæ . f.</i> Afrika , ke .
 f.◊<x>Afrika , i Poludne vighjascèse , ne beş
 suşaa◊priklopjeno oruşcjem , i polùvjerstvom od
 Van-|dalaa .</x> Gjør. S. Ben.

Agguagliamento , cioè l' agguagliare . <i>Æqua-
 <cb>tio , nis . f.</i> Jednacegne , gna . n.
 Takme-|gne , gna . n. Iştakmegne , gna . n.
 Işjed-|nacegne , gna . n.

Alato . add. che ha ale . <i>Alatus , a , um
 .</i>◊Krilat , ta , to . Krilaft , fta , fto .
 Krilataft ,◊fta , fto . Letufcti , fcta , fcto .
 Letufcte ,◊fcta . n. fuft .◊<x>I gnom' platta
 svud letuscta◊Po planini varti , i vjie .</x>
 Gjør. Raş. Chi◊ha il piede alato , o veloce .

<i>Alipes , dis .</i>◊Krilònogh , noga , nogo
.<cb>

Allegatore , chi allega , chi cita . <i>Qui afferi
.</i>◊Prikaçcjujuchi , uchja , uchje . Prikaçalaz
,◊aoza . m. Işvoditegl , glja . n.

Allegro , fuft. cioè l' uomo di natura al-|legro .
<i>Vir ingenio hilari .</i> Vefegljak , aka .
m.◊<x>Gljubàvje , şcivjeti s' druşcinom , junazi
,◊Rujno vinze piti s' dobrjem vesegliazi
,</x>◊Tir. Vefeles naravi . Vefeles chjudi .

Allentare , far lento . <i>Laxo , xas .</i>
Popufctati ,◊tavam , tiòfam . Popuftiti , vam ,
ftiòfam .◊Oslabiti , vam , biòfam . Opruşciti ,
şci-|vam , şciòfam . Odpuftiti , vam , ftiòfam
◊v. g. Odpuftavam paas . Allento la cinta
◊Allentare troppo . <i>Nimis relaxare .</i>
Raşla-|biti , slabivam , biòfam . Raşpuftiti ,
pufti-|vam , tiòfam . Raşpruşciti , şcivam ,
şciò-|fam . Allentare poco . <i>Parum laxare
.</i> Po-|puftiti , pufctavam , tiòfam .
Popruşciti ,◊şcjujem , şciòfam .

Alternare , cioè far una cofa a vicenda . <i>Al-
|terno , as , avi , atum .</i> Işmjentife , gnu-
|jèmfe , niòsàmfe . Promjentife , gnujèm-|fe ,
niòsàmfe . Ciniti fctògod na işmjenu .◊v. g.
Pripovjedati na işmjenu .◊<x>Na işmjenu kî
pritvarasc◊Noch u svjetlos dan u tmine ,</x>
Gjor. Raş.◊Işmjentife u pripovjedagnu . Freq.
Işmje-|nivatife . Promjegnivatife .<cb>

Alzare , inalzare . <i>Effero , ers .</i> Uşdighnuti
 , ðişcem , dighnuòfam . Uşpeti , uşpignem
 , ðuşpeòfam . ð<x>Glasoviti nu pisaozi ðUşpescèjih
 na toliko , </x> Ofm. Podighnnti , ðişcem ,
 dighnuòfam . Uşvifiti , fcjujem , ðfiòfam
 . ð<x>Şcena oholà ne misli ino , ðNego uşvisit sebe
 , i sina .</x> Ofm. Pripeti , ðpignam , peòfam
 . ð<x>Kî Mustafu opet şcele ðNa sto zarski svoj
 pripeti .</x> Ofm. Pro-|peti , nem , peòfam
 . <cb><x>Şàsve da od gne , kad oholu ðGlavu
 iştakmit k' Viscgnem' , prope , ðStarmoglàvse
 buscj dolu ðVjecnjeh ogagn' u potope .</x> Ofm.
 Propi-|gnati , gnem , pignòfam . Pegnati , gnem
 , ðgnaòfam . ð<x>Cjem şlòsardi şmaj na mene glavu
 pegne , </x> ðGjor. Pfal. Podignuti , dişcem ,
 dignuòfam . ðDignuti , dişcem , nuòfam . Vifiti ,
 vifim , ðvifiòfam . ð<x>Nakitjeno dubjem celo ðPut
 nebesaa vise , i steru .</x> Palm. Chr. ðV.
 Inalzare . Alzar in alto . <i>In fublime ðerigere
 .</i> Uşdignuti , nivam , nuòfam . Dig-|nuti ,
 dişcem , nuòşam . ð<x>Ter dignite , ter
 uprâvte ðVascju pamet put nebesaa .</x> Palm.
 Chr. ðAlzare le mani al cielo . <i>Tollere manus
 ad ðcælum .</i> Uşdignuti ruke , k' nebu .
 Uproftiti ðrukee k' nebu . ð<x>Tim put dvora
 nebeskoga ðOcci uprje , rukee uprostj .</x> Palm.
 Chr. ðK' nebu pruşciti rukee , pruşciam , şciòfam
 . ðAlzare di fotto . <i>Elevare .</i> Podignuti .
 Po-|diti , illi dignuti . Podpodignuti , dişcem

,Øgaòfam . Alzar qualcheduno . <i>Erigere ali-
|quem .</i> Uşdignuti , alliti podignuti tkoga
.ØAlzare la mano per percuotere .
<i>ElevareØmanum ad percutiendum .</i> Şamahnuti
, mah-|nujem , nuòfam . Freq. Şamahivati , hivam
,Øvòfam . Alzare da terra qualche cofa
.Ø<i>Elevare .</i> Dokuciti s' tle kojùgod ftvar
,Øcjujem , ciòfam . Freq. Dokucivati , civam
,Øviòfam . Alzarfi in punta di piedi . <i>Confi-
|ftere digitis .</i> Propetife na parfte , propi-
|gnèmfe , peòsàmfe . Ifpetife , pignàmfe
,Øpeòsàmfe . Freq. Işpignatife , gnàmfe , gnaò-
|sàmfe . Pripetife na parfte , pripignèmfe
,Øpripeòsàmfe . Alzarfi da letto . <i>Surgere
eØlecto .</i> Uftati , uftajem , ftaòfam .
Dignutife ,Øalliti uftajati iş odra . Alzarfi in
piedi . <i>Af-|furgo , gis .</i> Dignutife ,
dişcèmfe , nuòsàm-|fe . Uftati na noghee .
Podvignutife , vi-|şcèmfe , nuòsàmfe . Uftati ,
tajem , taòfam .Ø<x>A Dilaver dogovorj ,ØParvi
Veşier parvi usta .</x> Ofm. Alzarfi .Ø<i>Surgere
.</i> Uşdignutife , dişcèmfe , gnuòsàm-|fe .
Uftati , ajem , aòfam . Digntitife , di-|şcèmfe ,
dvignuòsàmfe . Podignutife , şcèm-|fe , nuòsàmfe
.Ø<x>Şa tvojùchje vechju szjennuØZarkve podvignut
neişbrojene .</x> Pal. Chr.ØUşraftjeti , illi
uşraftiti .Ø<x>Do zarskoga isctom stola ,ØOdkli
pade , da uşrasti .</x> Ofm. Chi alza
.<pb><x>Erigenis ,tis .</x> Uşdignitegl , glja .

m. Podi-|gnitegl , glja . m. Podiŝcjuchi , chja ,
chje ;◊e il fem. Podignitegliza , ze . f. Uŝdi-
|gnitegliza , ze . f.

Ambafciata , quel che riferifce l' ambafciado-|re .
<i>Mandatum , ti . n.</i> Poruka , ke . f. Far◊l'
ambafciata . <i>Mandata exponere .</i> Prikaŝa-
|ti poruku . Opovidjeti poruku . Portare
l'◊ambafciata . <i>Mandata ad aliquem afferre
</i>◊Donjeti tkomu poruku . Mandar ambafcia-|ta
ad uno per un altro . <i>Mandata alicui◊ad
aliquem dare .</i> Poruciti fctògod tkomu◊po
tkomùgod .

Amello , erba . <i>Amellus .</i> Divglifciza , ize
.◊f. Rutta divja , vje . f.

Ammonimento , ammonizione . <i>Admonitio ,◊nis .
f.</i> Nauk , ka . m. Svjetovagne , gna .◊n.
Uvjecbagne , gna . n. Nagovaragne ,◊gna . n.
Napomenutje , tja . n. Ofvjestje-|gne , gna . n.
Pofvjestjegne , gna . n. Svje-|ftjegne , gna . n.
Anelletti , a' quali fi affibbiano gli
uncinelli◊delle vefti . <i>Anuli veftiarii .</i>
Spone , naa .◊f. plur.

Anello , che fi porta in dito . <i>Anulus , li .
m.</i>◊Parften , na . m. Anello fenza pietra .
<i>Anu-|lus purus .</i> Parften beŝ ofcza .
Anello fpo-|falizio . <i>Anulus pronubus .</i>
Rukovni par-|ften. Parften blagofovjeni . Ŝarucni
par-|ften. Ŝcenidbeni parften. Parften pirni
.◊<x>I prid bojnom sabgljom gljutom◊Parsten pirni

cjùse uljesti .</x> Ofm. Anelloðd' oro .
 <i>Anulus aureus .</i> Šlatni parften.ðLuogo , o
 parte dell' anello , in cui s' in-|castra la
 gemma . <i>Pala , æ . f.</i> Bok odðparftena .
 Işnofito od parftena . Anello in-|gemma , che
 ha la gemma . <i>Anulus gem-|matus .</i> Parften
 s' ofczem , s' draghjem ka-|menom . Anello da
 figillare . <i>Anulus fi-|gnatorius .</i>
 Pecjatni parften. Parften pe-|cjatnik . Anello da
 cucire . V. Ditale .ðAnello , o anelletto da
 cortina . <i>Anuliðcortinales .</i> Biocjuh ,
 cjuha . m. Anello daðcatena . <i>Catenæ anulus
 .</i> Biocjuh , ha . m.ðPorfi , o metterfi l'
 anello . <i>Anulum indue-|re .</i> Staviti
 parften na ruku . Parften na-<pb>taknuti , ticem
 , nuòfam , illi natakòfam .ðLevarfi l' anello .
 <i>Deponere anulum .</i> Ski-|nuti , kinivam ,
 nuòfam . Staknuti , kni-|vam , knuòfam . Smaknuti
 , knivam , knuò-|fam parften s' ruke . Portar l'
 anello inðdito . <i>Digito anulum geftare .</i>
 Nofiti , nofim ,ðnofiòfam parften na ruzi . Fare
 , o lavo-|rare anelli . <i>Fabricare anulos .</i>
 Kovati , ku-|jem , kovòfam parftene .
 Anguillaia , luogo , o ridotto di anguille
 .ð<i>Anguillarum receptaculum .</i> Jeguglifcte
 ,ðfcta . n. Mjefto jegugljam' plodno .
 Annebbiamento , lo annebbiarfi . <i>Nebula te-|gi
 .</i> Omaghlegne , gna . n. Namaghlegne ,ðgna .
 n.

Anno . <i>Annus , ni . m.</i> Godifcte , fcta .
n.◊Godina , ne . f.◊<x>Evo gre pet godin' , dàsam
gne sluga ja ;</x>◊Scifc. Ljetto , ta . n. Anno
nuovo . <i>Annus◊incipiens .</i> Mlado ljetto .
Novo godifcte .◊Un' anno . <i>Annus , i . m.</i>
Godifcte danaa .◊Jedno godifcte . Mezzo anno .
<i>Semiannus ,◊ni . m.</i> Po godifcta . indecl .
Po ljetta . Po-|loviza godifcta . Un' anno , e
mezzo . <i>Sef-|quiannus , ni . m.</i> Godifcte i
po . Ljetto i◊po . D' anno in anno . <i>In annos
fingulos .</i>◊Od godifcta do godifcta . Od
ljetta do ljet-|ta . D' un' anno . <i>Annulus ,
a , um .</i> Od◊jednoga godifcta . Od jednoga
ljetta . Jed-|nòljetni , na , no . Godifctni , na
, no ,◊illi od godifcta . Jednògodni , na , no
.◊Jednògodaz , dza . m. Di due anni . <i>Bi-|mus
, a , um ,</i> Dvòljetni , na , no . Dvò-|godni ,
na , no . Dvògodaz , dza . m. Od◊dva godifctaa .
Di tre anni . <i>Trimus , a ,◊um .</i> Tròljetni
, tna , tno . Trechjak , aka .◊m . Di quattro
anni . <i>Quadrinus , a , um .</i>◊Cetveròljetni
, na , no . Cetverògodni ,◊Cetverògodaz , dza .
m. Cetvartak , aka .◊m. fuft. Di cinque anni .
<i>Quinquennis . &◊hoc quinquenne .</i>
Peteròljetni , na , no .◊Peterògodni , na , no .
Peterògodaz , dza .◊Petak , aka . m. fuft. Di fei
anni . <i>Se-|xennis , fexenne .</i>
Scefteròljetni , na , no .◊Scefterògodni , na ,
no . Scefterògodaz ,◊dza . m. Sceftak , aka . m.

fuft. Di fette◊anni . <i>Septennis . & hoc ne
 .</i> Sedmeròljet-|ni , tna , tno . Sedmerògodni
 &c. Sedmak ,◊aka . m. Di otto anni . <i>Octennis
 . & hoc◊ne .</i> Ofmeròljetni , tna , tno .
 Ofmerògodni◊&c. Ofmak , aka . m. fuft. Di nove
 anni .◊<i>Novennis , & hoc ne .</i>
 Deveteròljetni , na ,◊no . Deveterògodni &c.
 Devetak , aka .◊m . Deveterògodaz , dza . m. Di
 dieci an-|ni . <i>Decennis , & hoc ne .</i>
 Defeteròljetni ,◊na , no . Defeterògodni , na ,
 no . Defe-|tak , aka . m. fuft. Due anni .
 <i>Biennium ,◊ii . n.</i> Dva godifcta . Dvie
 godine danaa .◊Dva ljetta . Tre anni , tempo di
 tre an-|ni . <i>Triennium , nii .n.</i> Tri
 godifcta . Tri◊godine danaa . Tri ljetta . Tre
 anni com-|piti , o finiti . <i>Trieteris plena
 .</i> Tri punna◊godifcta . Tri zjela godifcta .
 Quattro an-<cb>ni , lo fpazio di quattro anni .
 <i>Quadrien-|nium , nii . n.</i> Cetiri godifcta
 . Cetiri go-|dine danaa . Cinque anni .
 <i>Quinquennium ,◊nii . n.</i> Pet godifctaa .
 Pet ljetaa . Pet go-|dinaa .◊<x>Evo gre pet
 godin' . . .</x> Scif. Che dura cin-|que anni .
 <i>Quinquennalis , le .</i> Petòljetni ,◊na , no
 . Sei anni , lo fpazio di fei anni .◊<i>Sexennium
 , nii . n.</i> Sces godifctaa . Sette◊anni .
 <i>Septennium , nii . n.</i> Sedam godi-|fctaa .
 Sedmòljetni , na , no . Otto anni .◊<i>Octennium
 , nii . n.</i> Osam godifctaa . Ofmò-|ljetni , na

, no . Nove anni . *<i>Novennium* ,*ñii* . n.*</i>*
 Devet godifctaa . Devetòljetni ,*ñna* , no . Dieci
 anni . *<i>Decennium* , *nii* . n.*</i>*◊Defet
 godifctaa . Defetòljetni , na , no .◊Molti anni .
<i>Multi anni .*</i>* Mnoga godifcta .◊Velika
 godifcta .◊*<x>*Ni mene raşdruşcit' godiscta
 velika◊Nèmogu , da sluşcit' ne buddem do vjika
 .*</x>*◊Bur . Molti anni fa . *<i>Multis* ab hinc
 annis .*</i>*◊Mnogo godifctaa na şada . Jeft mnogo
 go-|difctaa . Davno . Sono due anni .
<i>Duobus◊ab hinc annis .*</i>* Jefu dua godifcta .
 Na şad◊dua godifcta . Di quì a dieci anni .
<i>Ad◊decem annos .*</i>* Do defet godifctaa . Şa
 defet◊godifctaa . Nakon deset godifctaa . Ove de-
 |fet godifctaa . Per un' anno . *<i>In* annum
 .*</i>*◊şa godifcte . Per l' anno fequente . *<i>In*
 pro-|ximum annum .*</i>* Şa do ljetta . Şa do godi-
 |fcta . Şa dohodno godifcte . Anno bifefti-|le ,
 o del bifefto . *<i>Annus* intercalaris .*</i>* Pri-
 |ftupno godifcte . Priftupno ljetto . Anno◊Santo
 , ovvero , anno del Giubileo . *<i>An-*|nus
 fæcularis .*</i>* Sveto godifcte . Anno lu-|nare ,
 quale dura un mefe . *<i>Annus* luna-|ris .*</i>*
 Mjefecno godifcte . Anno folare . *<i>An-*|nus
 folaris .*</i>* Suncjano godifcte . Godifcte◊od
 dvanes mjesezaa . Di molti anni , co-|lui ; che
 ha molti anni . *<i>Annofus* , a , um
 .*</i>*◊Davgnigni , na , no . Davgni , gna , gne
 .◊*<x>*Kjêm sred viecne pechi davgne◊şlobne duscee

muce , i parşce .</x> Palm. Chr.◊Pun godifctaa .
 Vremenit , ta , to .◊<x>Jürje muşc vremenit , a
 pjesni pripjeva ;</x>◊Ekt. Raş. Una volta all'
 anno . <i>Semel in◊anno .</i> Jednòkrat na
 godifcte . Jedan put◊na godifcte . Jednom na
 godifcte . Finito◊l' anno . <i>Anno exacto .</i>
 Na işhod godifcta .◊Na fvarhu godifcta . Un' anno
 sì , e l' al-|tro no . <i>Alternis annis .</i>
 Svako drugo go-|difcte . Svaka dva ljetta .
 Dentro al ter-|mine d' un' anno . <i>Intra annum
 .</i> U jedno◊godifcte . Più d' un' anno .
 <i>Ultra annum◊spatium .</i> Viſce od godifcta .
 Boglje od go-<pb>difcta . Vechje od godifcta . L'
 anno ap-|preſſo . <i>Poſt annum .</i> Drugo
 godifcte . Do◊hodgnega godifcta . Do godifcta .
 Do ljet-|ta . L' anno paſſato . <i>Anno præterito
 .</i> La-|ni . Lanko godifcte . Proſcjafſto
 godifcte ,◊alli ljetto . L' anno venturo .
 <i>Anno futu-|ro .</i> Dohodno godifcte . Due
 anni fono .◊<i>Duobus ab hinc annis .</i>
 Onomlani . Jefu◊dvje godine danaa . Jefu dva
 ljetta . Na◊fad dva godifctaa . Anno temperato .
 <i>An-|nus temperatus .</i> Meko godifcte . Anno
 in-|falubre . <i>Annus infalubris .</i> Nèſdravo
 go-|difcte . Anni della difcrezione , fono quel-
 |li , ne' quali uno difcerne il bene dal ma-|le .
 <i>Sapientitiæ anni .</i> Raşloşcna godifctaa
 .◊Spametna godifctaa . Raşborna godifctaa .◊Avere
 gl' anni della difcrezione , ovvero◊l' ufo della

ragione . *<i>Sapio , is , pui , itum .</i>*◊Imati
 raşborna godifctaa . Mochi raşabrati◊şlo od dobra
 . Biti ftupio na ljetaa od◊raşbora . Anni
 giovanili , o della gioven-|tù . *<i>Anni*
crefcentes .</i> Mlada godifctaa .◊Zvjat mladofti
 . Prolitje mladofti . Metaf.◊Proljetno doba ,
 illi proljetna .◊*<x>*Mladaz , tkomu fmart
 nèmilla◊U proljetno bjesce doba◊Priko reda dni
 skratila .*</x>* Palm. Chr. An-|ni della vecchiezza
 . *<i>Anni fenefcentes .</i>* Sta-|ra godifctaa .
 Anni della decrepitezza .◊*<i>Anni Acheruntici*
.</i> Vifcègrebna godifctaa .◊Najposledgna
 godifctaa . Prištara godifctaa .◊*<x>*Akòsam
 pristara , allìsam pristala ,*</x>* Gun.◊Dubr. Anni
 della puerizia . *<i>Anni pueriles .</i>*◊Djetinska
 ljettaa . Djetigna godifctaa . Neş-|rela doba .

Anfiare , ed anfare , respirare con difficoltà
 .◊*<i>Anhelo , as .</i>* Şapjehatife , pjehàmfe ,
 hòsàm-|fe . Trudnòdihati , diham , hòfam . Pjeha-
 |ti , ham , hòfam . Stegnati , ftegnem , fte-
 |gnaòfam .

Aratro di più paja di buoi . *<i>Aratrum mul-|tiplici*
jugo .</i> Plugh , ga . m.◊*<x>*Er scto uşore şa
 dan plugom ,◊S' opchinòmùje hrane dosta .*</x>*
 Ofm. Cavar◊qualche cofa coll' aratro . *<i>Arando*
effodere .</i>◊Işorati , rajem , raòfam .
 Işkopati , alli◊işnjeti ralom fctògodi . Cavato
 coll' aratro .◊*<i>Aratro effoßus .</i>* Işoran ,
 ana , ano . Il ferro◊dell' aratro , cioè il

vomero . <i>Vomer , ris .</i>◊Lemifc , fcja . m.
 Raonik , ika . m. Orno◊gvoşdje . Il dentale dell'
 aratro . Plaş , şa .◊m . Manico dell' aratro .
 <i>Aratri manubrium .</i>◊Raliza , ze . f.
 Darşcjak od rala .

Arbitrato , add. cofa giudicata dall' Arbitro
 .◊<i>Arbitratus , ta , tum .</i> Sudjen , na , no
 od◊dobrjeh gljudji .

Arcajo , ed arcaro , facitor d' arche . <i>Arca-|rius
 , rii .</i> Skrignar , ara . m.

Arrollare , fcrivere a rullo . <i>Confcribo , bis
 .</i>◊şapisati , fcjujem , faòfam . v. g.
 Vojnike◊&c. şbirati vojnike . Skupiti vojnike .
 Is-|pifati viteşove .<cb><x>K' jagnicjarim' ter
 savisce◊Vele vechi broj iş nova◊Po işboru josc
 ispisce◊Od Istocnjeh Viteşovaa .</x> Ofm.
 İşbirati◊vojnike .◊<x>A i cette
 Spahogljanaa◊Dvàkrat vechje s' tiem ucinj◊S' nova
 skupa işabrana .</x> Ofm. Arrollarfi .◊<i>Dare
 nomen militiæ .</i> Şapitatife şa vojnika
 .◊Upifatife u vojsku .

Affaffinato . <i>Aprædatoribus defpoliatus .</i> Po-
 |robgljen , na , no . Udart od gufe .

Attivo , che fà , o può fare . <i>Actuofus , fa ,◊fum
 .</i> Pofpjefcjān u djellu , na , no
 .<pb><x>Mnoşc na svako şlo pospjescna .</x>
 Palm.Chr.◊Obartan , tna , tno . Djellòtvorni , na
 , no .◊Otresen , ena , eno . Hitar . Lak . Barş
 u◊poslieh , u djellieh &c.

Aumentare , aumentarfi . V. Accrefcere , ac-
|crefcerfi .

Avoltóio , uccello di rapina . <i>Vultur , ris .
n.</i>◊Sep , pa . m. Koftòşcer , era . m.
Koftòşc-|der , ra . m. Dobròhran , na .
m.◊<x>Sluşci şa dom , i obranu◊şcderglivomu
Dobròhranu .</x> Gjorg. Pf. Di◊avoltoio .
<i>Vulturinus , a , um .</i> Koftòşcer-|ni , na ,
no . Koftoşcderni , na , no .

Avvenire , avv. per l' avvenire . <i>Posthac
.</i>◊Unaprjeda . Od fada .

Bámbagino , tela di bambagio . <i>Textile Goffi-|pium
.</i> Tkani bumbak . Pamukni poftav .◊Di bambagio
, <i>Goffipinus , a , um .</i> Bum-|bacni , na ,
no . Illir. Barb. Pamukni , na ,◊no . ufato .

Banco , ftromento di legno da federci più◊perfone
infieme . <i>Subfellium , lii . n.</i> Klup ,◊upa
, m. Sjedalifcte , fcta . n. Sjedifcte ,◊fcta .
n. Banco di galera , o altro naviglio
◊<i>Tranftrum , ftri . n.</i> Plavski klup .
Klup◊od voşazaa . Illir. barb. klup od ghjemie
◊Banco da due . <i>Bifellium , lii . n.</i> Klup
şa◊dvoizu . Stoo şa dvoize . Banco da Macel-|lajo
 . <i>Menfa lanionia .</i> Trup , pa . m. Tru-
|pina , ne . f.

Banda , cioè parte , o lato . <i>Pars , tis .
f.</i>◊Strana , ne . f.◊<x>Alli Dervisc desnu
stranu◊Hitro uşmace</x> Ofm.
Banda◊dinnanzi . <i>Pars antica .</i> Spredgna

ftrana .◊Sprjed , da . n. da ufarsi raramente .
 v. g.◊Moje fprjed ; in neutro . Od ovoga fprje-
 |da . Dalla banda dinnanzi . <i>A fronte .</i>
 Sprje-|da . S' fprjedgne ftrane . Banda di dietro
 .◊<i>Pars poftica .</i> Šadgna ftrana . Šad , od
 šada .◊n. fi ufa però di rado . v. g. Moje šad
 .◊Dalla banda di dietro . <i>A tergo .</i> Šada ,
 s'◊šadgne ftrane . A banda finiftra .
 <i>Siniftror-|fum .</i> Na ljevu ftranu . K'
 ljevoj ftrani . Na◊ljevu . A banda deftra .
 <i>Dextrorfum .</i> Na<pb>defnu ftranu . K'
 defnoj ftrani . Na defnu .◊Banda di dentro .
 <i>Pars interior .</i> Unutar-|gna ftrana .
 Unutra . v. g. Moje unutra ;◊in neutro . Banda di
 fuori . <i>Pars exterior .</i>◊Išvagna ftrana .
 Dvorgna ftrana . Šdvora .◊v. g. Moje šdvora .
 Banda di fotto . <i>Pars◊inferior .</i> Dogna
 ftrana . Strana odišdo ,◊alliti ošdala . Banda di
 fopra . <i>Pars fupe-|rior .</i> Gorgna ftrana .
 Strana od šgori , al-|liti odišgara . Da qualche
 banda . <i>Alicun-|de .</i> Od kègod ftrane . Od
 kudàgodi . Od◊klègodi . Da ogni banda .
 <i>Undique .</i> Od◊fvake ftrane . Od fvjeh
 ftranaa . Sa fvjeh◊ftranaa . Odàfvuda . Od
 fvudjere .◊<x>A to , da tuj ne pašechi◊Sipat
 pjeneš od svudjere .</x> Ofm. Da ban-|da a banda
 , come paffar da banda a ban-|ba .
 <i>Tranfverbero , as , avi , atum .</i>
 Probiti◊s' ftrane na ftranu , biam , biòfam

.*<x>*Pribjenèsu cjavlìm k' darvu*>*S' ove strane na
onu stranu .*</x>* Mand. Bun.*>*Prodjefti tkoga
kopjem , macem &c. dje-|vam , diòfam
.*>**<x>*Sprjèdnu işide scip proş plechja .*</x>*
Ofm.*>*Da banda , da parte . *<i>*Seorfim .*</i>* Pò
nafe .*>*S' ftrane . Na ftranu . Na pònafe . Na pò-
|fe . Pòfe . Banda , traverfa di drap-|po , che
fi porta al colo . *<i>*Fafcia collaris
.*</i>**>*Nadrameniza , ze . f. Prikorameniza , ze
.*>*f. Banda , compagna de foldati . *<i>*Cohors
,*>*tis . f.*</i>* Cetta , te . f. Banda di cavalli ,
o*>*di foldati a cavallo . *<i>*Turma equitum .*</i>*
Cet-|ta kognikaa , illiti kognicka cetta .
Da*>*ogni banda , da ogni lato , da ogni parte
.*>**<i>*Quoquoverfus .*</i>* Od fvjeh krajaa . Od
fvjeh*>*ftranaa . Oda fvuda .

Beneficio , grazia , piacere . *<i>*Beneficium , cii
.n.*</i>**>*Gljubav , vi . f. Millos , fti . f. Har ,
ri . f.*>**<x>*Er szjenech' , da şa har imatichje
scto cinih .*</x>**>*Scifc. Dobròdarje , rja . n.
Dobròcinfivo ,*>*ftva , n. Render beneficio .
*<i>*Beneficium red-|dere .*</i>* Vratiti gljubav .
Vratiti har , alliti*>*millos . Beneficio , entrata
, o rendita cle-|ricale . *<i>*Vectigal
Ecclefiasticum .*</i>* Zarkovni do-|hodak . Nadarje
Zarkovno . Djacja Zarkov-|na . Beneficio femplice
. *<i>*Sacerdotium curio-|natu vacuum .*</i>* T.
Eccl. Zarkovni dohodak*>*beş pomgne od dufcjaa .
Slobodna djacja .*>*Beneficio curato , che ha cura

d' Anime .*<Sacerdotium Animarum cura affectum*
. T.*&Eccl.* Zarkovno nadarje s' pomgnom
od*&dufcjaa* . Djacja s' vladagnem od Dufcjaa
&djacja dufcjòvladna , illi Vladòdufcja . Be-
|neficio di molta rendita . *<Sacerdotium pin-*
|gue *.* T. Eccl Nadarje Zarkovnu bogato
&Djacja bogata . Beneficio piccolo , di
poca*&rendita* . *<Sacerdotium tenue* *.* T.
Eccl. Nada-|rje Zarkovno mallòkorifno .
Mallòkorisna*&Djacja* . Uboga Djacja .

Bigamía . *<Bigamia* , mie *.* Dvòšcegnegne , gna .
n.

Borráccia , fiafco di cuojo da tener vino , o*&altro*
liquore . *<Scortea* , teæ .f.*<Buklia* , e
.f. Matara , re . f.

Bottigliere , colui , che tiene cura del vino ,*&che*
fi dà a tavola . *<Vini epularis curator*
*.**&Nadpivnik* , ka . m. Kgljucjar od vinize
&Kuchni karcmar .*&x>Kgljucèmi oncjas da od*
karcme Bolkovich :*&Kruh* , larad , vladam svîm , a
kupgljùse*&u* vinu *.**<x>* Darşc. Raşl.

Broccoli , cime di cavoli più tenere , e deli-|cate .
<Brafficarum cymæ . f. pl *.* Prokoliza ,*&ze*
. f. Klize kupufne . f. pl. Prokuliza .

Bufalo . *<Bubalus* , li . m.*<Bivoo* , la .
m.*&x>Bivoli* od sprjed , kjî do
pasa*&Cjovjecjanske slike nose* *.**<x>* Ofm. Carne
di*&bufalo* . *<Caro bubalina* *.* Bivoluie mefo
. Bi-|volsko mefo .

Bugiare , dir bugia , mentire . <i>Mentior , iris
. </i>◊Lagati , laşcem , lagaòfam . Slagati ,
laşcem ,◊slagaòfam . Svoditi laşc , fvodim ,
fvodiò-|fam laşc .◊<x>Njednùt' laşc ne svodim
&c.</x> Elekt.<cb>

Capria , capivola , capra falvatica . <i>Caprea ,◊æ .
f.</i> Sarna , ne . f.

Cefto . V. Caneftro .

Chiarire , render chiaro cofa , o negozio
.◊<i>Dilucide explicare .</i> Raşabrati , birem ,
braò-|fam . Raşbiftriti , trivam , triòfam .
Occito◊prikaşati , &c. Chiarire uno , dicefi di
chi◊fa qualche burla ad un altro . <i>Imponere a-
|licui .</i> Skroiti tkomu , skroim , skroiòfam
.◊Metaf. Chiarirfi di qualche cofa .
<i>Certum◊aliquid fcire .</i> Dofvjeftitife ,
tujèmfe , tiò-|sàmfe . Dochi u poşagne . V.
Avvederfi .

Chirógrafo . <i>Chirographus .</i> Rukòpifagne , gna
.◊n. Rukòpis , fa . m.

Chiunque . V. Qualunque .<pb>

Coagulare . V. Quagliare .

Coeguale , add. in fieme eguale . <i>Coaqualis ,◊& hoc
le . n.</i> Jednòvarftan , na , no .

Coetanea . <i>Coæva , æ .</i> Varfniza , ze . f.

Cóltrice , feconda breve . V. Matarazzo .

Commettere , dar commiffione . <i>Demando , as
. </i>◊Narediti , djujem , diòfam . Odrediti ,
dju-|jem , diòfam . Priporuciti , cjujem , ciòfam

.◊Nasloniti tkomu , naslagnam , niòfam . Şa-
|povidjeti , dam , diòfam . Freq. Narediva-|ti .
Priporucivati . Naslonivati tkomu pofo
.◊Commettere errore . V. Errare . Commet-|tere ,
fare . V. Fare . Commettere , inca-|ftrare . V.
Incaftrare .

Cómputo . V. Abbaco .

Comunque , in qualunque modo . <i>Utcumque
.</i>◊Kakòmùdrago . Na koìmùdrago nacin . Co-
|munque fia . <i>Ut ut eft .</i> Kakòmùdrago .

Cóncia , luogo , dove fi conciano le pelli
.◊<i>Coriarii officina .</i> Tabakaria , rie . f.

Conclufióne , termine del difcorfo . <i>Conclu-|fio ,
nis . f.</i> Şaglava , ve , f.<pb><x>Na şaglavu
tjeh besjedaa &c.</x> Palm. Chr.◊Şaglavgljegne ,
gna . n. Svarha , he . f. Do◊varfcegne , gna . n.
Şavarfcja , e . f. Con-|clufione , dicefi ad ogni
propofizione di◊quelle , che fi pubblicano per
difputare .◊<i>Pofitio , nis . f.</i> Şaglavak ,
vka . m. Soften-|tar , o difender la conclufione
 . <i>Thefer propu-|gnare .</i> Braniti şaglavke .
Softentor , o di-|fenfor delle conclufioni .
<i>Propugnator the-|fium .</i> Branitegl
şaglavakaa .

Conquaffatrice . <i>Conquaffatrix , cis . f.</i>
Raşru-|fcitegliza , ze . f. Satritegliza , ze .
f. Sa-|tarifaliza , ze . f.

Contrário , fuft. <i>Vír adverfus .</i> Nepriategl
,◊glja . m. Protivnik , ka . m. Suprotivnik ,◊ka

. m. Per lo contrario , avv. ovvero all'◊incontro
. <i>Contra .</i> S' drughe ftrane .

Convertíto . <i>Converfus , a , um .</i> Obrachjen
,◊na , no . Priveden na dobar put . Obarnut ,◊ta
, to . Svarnut , ta , to .

Corallo , pianta , che nafce nel fondo del◊mare .
<i>Corallium , lii . n</i> Koragl , glja .
m.◊<x>Od koraglja usti objavj ,◊A od lira parsi
svoje ,</x> Rařm. Di corallo .◊<i>Corallinus , a
, um .</i> Koraglni , na , no .◊<x>Koraglne dvje
usti &c.</x> Gljub.◊<x>Proř koraglana neka
vrata◊Na dvor řcjarko Sunze iřide .</x> Ofm.

Corbacchiotto , corbicino , il figlio del cor-|vo .
<i>Corvi pullus .</i> Gravranich , chja .
m.◊Vranich , chja . m.

Cordiale , cioè intimo . <i>Intimus , a , um
.</i>◊Sarcjani , na , no .

Corrente , fuft. acqua , che corre . V. Cor-|fia .
Corrente , furiofo . <i>Torrens , tis .
m.</i>◊Poplaviza , ze . f.◊<x>Tako kàdse niř
brjegh starmi◊Poplaviza plaha obarne .</x> Palm.
Chr. Con◊la corrente , fecondo la corrente .
<i>Secundo◊flumine .</i> S' vodom . Niř vodu .
Contro la◊corrente , contro l' acqua . <i>Amne
contrario .</i>◊Suproch vodi . Uř vodu .

Corvo . V. Corbo .

Cofcienza . V. Confcienza .

Coftipáto nel detto fenfo . <i>Alvus conftipatus
.</i>◊Tjelo řatvoreno .

Cótica , pelle dura dell' animale . V. Cuojo .
 Cozzóne , colui che doma i cavalli .
*<i>Domitor*equorum .*</i>* Kognicjar , ara . m.
 Credo , il fimbolo degli Apoftoli , o della*<i>*Fede .
<i>Apoftolorum fymbolum .</i> T. Eccl. Vje-
 |rovagne , gna . n.
 Crifto , Dio , ed uomo infieme , e Salvator*<i>*del mondo
 . *<i>Chrifthus , fti .</i>* Ifukars , Ifukar-|fta .
 m. Crifto crocififfo . *<i>Chrifthus cruce pen-
 |dens .</i>* Ifukars propeti .*<i>Defcr .</i>**<x>*Bjesce
 Spasitegl od svjeta rašpet , i rašteghnut*<i>*na
 krišcju , goo , nagh , išderan ,
 išraghen.*<i>*Tezjasce karv Bošcja s' glave po kosah
 . po*<i>*bradi , po svemu obrašu . Tochjahùse iš no-
 |gaa , iš rukaa potozi privrjedne karvi ,
 i*<i>*nèimasce ghdi nasloniti plemenitu
 glavu*<i>*dracjami okrugnenu .*</i>**<x>* Kafs. If. Crifto
 rifu-|fcitato . *<i>Chrifthus redivivus .</i>*
 Ifukars uskarf-|nut .
 Crocetta , croce piccola . *<i>Crux parva .</i>* Kri-
 |ščjak , šcka . m. Kriščich , šcichja . m.
 Croftáta , forte di torta , che fi fa di frutti*<i>*cotti
 con zucchero tra fottiliffime crofte di*<i>*pane .
<i>Placenta cruftata .</i> Slatka kruftavi-|za .
 Prifnaz , od prifnaza . m. propriamen-|te torta
 col formaggio .
 Cubo , corpo quadrato di fci lati eguali . *<i>Cu-|bus
 , bi . m.</i>* Scefterònughla , ghle . f.

Curiófo , add. che ha curiofità . <i>Curiofus , a^{um}
 .</i> Şcegljan , eglna , eglno . Naghli , gla
 ,^oglo .^o<x>Cîm lipote draghe toli^oNaghlim okkom
 svaki kupi .</x> Ofm. Naglò-|şcegljan , glna ,
 glno . Curiofo , chi tien^ocura , o fi prende
 penfiero di qualche co-|fa . <i>Curiòfus , a , um
 .</i> Pomnechi , chja ,^ochje . Pomnik , ka . m.
 Poman , mna , mno .^oCuriofo di fapere .
 <i>Sciendi cupidus .</i> Şce-|glnòşnanaz , nza .
 m. Şcegljan şnati .

Danneggiatrice . <i>Damnifica .</i> Ofctetitegliza
 ,^oze . f.

Decína , quantità , o numero di dieci . <i>De-|cas ,
 dis . f.</i> Defetero . Defetak . Defetina ,^one .
 f. Mezza decina . <i>Pentas , dis . f.</i> Pe-
 |tero . Pet . Petina , ne . f.

Degradáre , privare della dignità facerdota-|le .
 <i>Sacerdotali gradu exuere .</i> Potifnuti
 s'^oRedovnictva . Obaliti s' vifine
 doftojanftva^oRedovnickoga .

Diamantino , fatto di diamante . <i>Adamani-|nus , a
 , um .</i> Od diamanta .

Dibattúto , difputato . <i>Difcuffus , a , um .</i>
 Pro-|trefen , na , no . Raşcefcgljan , na , no .

Diffidenza , difconfidenza . <i>Diffidentia , tiæ
 .^of.</i> Nèuşdagne , gna . n. Neufagne , gna .^on.
 Nèuşdanftvo , ftva . n. Nèpriaşnos , fti . f.

Dignità , cioè grado d' ufficio , o preeminen-|za .
 V. Degnità .

Dimagràre , fmagrire . V. Smagrare .

Dire , vale efprimere con parole . <i>Dico , cis .</i>◊Rechi , illi rjeti , govorim , rekaòfam . Go-|voriti , rim , riòfam . Praviti , vim , viòfam .◊<x>On şacjudjen pravi ; Tìli◊Sluşi tvomu peresc noghe ?</x> Palm. Chr.◊<x>I gnegov tac şcestok ogágnse svjem pravi ,</x>◊Ragn. Raşl. Klikovati , kujem , kovaòfam .◊<x>Pak u sebi klikovati◊Pocce ; ovòsu svete noghe ,</x> Mand. Bun.◊Veliti , lim , liòfam . Işuftiti , tjujem , tiòfam .◊<x>Jesus , i ove rjeci işustj .</x> Palm. Chr. Be-|sjediti , besjedim , diòfam .◊<x>Ter pogledim' veselime◊Tuşcnjem rjeci sej besjedi .</x> Palm. Chr.◊Işrizati , zam besjede , işrizòfam .◊<x>I şcjalosno işrizati◊Milloy druşcbi sej besjede .</x> Palm. Chr.◊Puftiti besjede .◊<x>I medenjeh kada ustji ,◊Iş kjêh rajska rados grede ,◊Put Nebesa Jesus pustj◊Vjecnom' Otzu sej besjede ;</x> Palm. Chr.◊Oglafitife u rjeci &c.◊<x>U ovakèse tuşcbe oglasj ,</x> Ofm. Dire , rac-|contare . V. Raccontare . Altro dire , al-|tro fentire , altrimenti penfare , altrimen-|ti parlare . <i>Alia loqui , alia fentire .</i> Jedna◊govoriti , druga chjutiti . Nèodgovarati je-|şik farzu . Dir tutto . <i>Totum dicere .</i> Işgo-|voriti , varam , riòfam . Işrjeti , işrecem ,◊işrekòfam . Dire brevemente . <i>Strictum di-|cere .</i> Rjeti u kratko . Rjeti u jednu rjec .◊U mallo

rjecji . Dire a bocca . <i>Coram lo-|qui .</i>
 Rechi na okkòrjec . Rechi na rjec .◊Dire a
 capriccio , parlare incoufideratamen-|te .
 <i>Inconfiderate loqui .</i> Govoriti fcto do-
 |ghje na ufta . Govoriti nèraşborito . Spar-|dati
 , dam , fpardaòfam . Cepgljuskati ,◊kam , kòfam .
 L' una , e l' altra . V. Bafs.◊V. Ciarlare . Dire
 a lettere di fcattola ,◊cioè chiaramente .
 <i>Clare dicere .</i> Oblo go-|voriti . Govoriti
 efckero . Occito rechi . Go-|voriti bjelòdano .
 Dire alla sfuggita . <i>Cur-|fim dicere .</i>
 Grede govoriti . Hode rechi◊fctògod . Uşgrede
 rjeti . Dire in fcreto .<cb><i>Secreto dicere
 .</i> Dorizati potajno , alli u◊potaj .◊<x>Jedan
 drugom' viteş s' parva .◊Ovi u potaj glas doriza
 .</x> Rechi potajno .◊Dir la bugia . Rjeti laşc .
 Lagati . Dir di◊sì . <i>Ajo , is .</i> Veliti ,
 lim , liòfam , alli rje-|ti takòje . Dir di no .
 <i>Nego , as .</i> Njekati ,◊kam , kaòfam .
 Şanjekati , kam , kòfam .◊Rjeti , njie tako &c.
 Dir francamente .◊<i>Præfenti animo dicere .</i>
 Slobodno govoriti .◊Dire all' improvvifo . <i>Ex
 tempore dicere .</i>◊Işnènade besjediti . Nènadno
 govoriti . Re-|chi nèmisliufci . Dir male d'
 alcuno . <i>Male-|dice de aliquo dicere .</i> Şlo
 o tkomu govori-|ti . Şlòrechi od tkoga , recem ,
 kaòfam .◊Kofiti tkoga jeşikom. V. Detrarre .
 Chi◊dice mal d' alcuno . <i>Obloquutor , ris .
 m.</i>◊Şlòrek , ka . m. Şlògovornik , ka . m.

Şlò-|govoraz , za . m. Il femm. Şlògovorniza , ze
 .◊Dir all' orecchio . <i>In aure dicere .</i>
 Pofcjap-|tati , fcjapchjem , taòfam . Scjaptati ,
 chjem ,◊taòfam . Govoriti na uho . Freq.
 Scjapchi-|vati , vam , vòfam . Pofcjapchivati ,
 vam ,◊vòfam . Dire a propofito . <i>Apte dicere
 .</i>◊Sputno govoriti . Skladno , podobno ,
 po◊putu govoriti . Dire a fpropofito .
 <i>Inepte◊dicere .</i> Nèfputno besjediti .
 Nèskladno , nè-|podobno govoriti . A dir il vero
 . <i>Ut ve-|rum dicam .</i> Şa rechi pravo . Şa
 rechi ifti-|nu . Dico il vero . <i>Verum dico
 .</i> Pravvo dim .◊Govorim iftinu . Dir di cuore
 . <i>Ex animo◊dicere .</i> Govoriti od farza .
 Govoriti iş far-|za . Dir da dovero . <i>Serio
 dicere .</i> Govoriti◊na sbigl . Dire a mente .
 <i>Dicere memoriter .</i>◊Rechi na pamet . Dire
 leggendo . <i>Dicere◊de fcripto .</i> Rechi iş
 kgnighe . Dir non vo-|lendo , dir una parola per
 un' altra . <i>Labi◊lingua .</i> Utechi Jeşik .
 Şarechife , ricèmfe ,◊şarekaòsàmfe . Dir l'
 ifteffo . <i>Cantilenam◊candem canere .</i>
 Pjevati iftu pjefan. Sve◊jedno govoriti . Dir pro
 , e contra a qual-|cuno . <i>In utramque partem
 dicere .</i> Şa tko-|ga , i na tkoga rechi . Dir
 la cofa come◊fta . <i>Rem ut eft eloqui .</i>
 Rjeti onako , kakò-|je . Dico a te . <i>Tecum
 loquor .</i> Tebi govo-|rim . S' tobom govorim .
 Quefto ti dico .◊<i>Hoc ajo .</i> Ovòti velim .

Ovo dim .*<x>Dim ,Gascpara Millostichja ;</x>*
 Ofm. Si dice .*<i>Fama eft .</i>* Glas jeft .
 Govorife . Per dir me-|glio . *<i>Ut verius dicam*
.</i> Şa boglje rjeti . Non*¶*fia detto per
 arroganza . *<i>Abfit verbo inju-|riae .</i>* Nèbud'
 receno şa okolaft .

Difcépolo . V. Scolaro .

Difordinatiffimo , ed inordinatiffimo , fom-|mamente
 difordinato . *<i>Inordinatiffimus ,¶ma , mum*
.</i> Prinèredni , na , no . Najnè-|rednji , njia
 , njie .

Diffigillabile , che fì può diffigillare . *<i>Refi-|gnabilis , & hoc le .</i>* Raşpecjatni , na , no
 .*¶*Odpecjatni , na , no .

Difverginäre . V. Sverginare .

Edile , Magiftrato fopra la vettovaglia , e an-|cora
 fopra glj edificj pubblici . *<i>Curator an-|nonæ*
.</i> Nadbrafcnik , ka . m. Nadljetinik ,*¶*ka .
 Vifce hrane , illi şgraghje .

Erbame . *<i>Herbæ . f. pl .</i>* Travje , vja .
 n.*<cb><x>Jestojskàje mâ jedina¶Nèpoşnano , garko*
travje .</x> Raşm.

Efplicäre , far intendere . *<i>Explico , cas .</i>*
 Ka-|şati , şcem , şaòfam . Tomaciti , cim , ciò-
 |fam .

Effere , nome , cioè ftato , effenza , efiften-|za .
<i>Status , us .</i> Bitje , tja . n.

Fáccia , volto , parte nota del corpo , *<i>Facies*
,¶ciei . f.</i> Lize , liza . n. Obraş , şa . m.

Faccia◊tonda . <i>Facies orbiculata .</i> Oblo
 lize . Fac-|cia puerile . <i>Species puerilis
 .</i> Djetinski o-|braş. Faccia di morto .
 <i>Facies cadaverofa .</i>◊Martacko lize . Faccia
 lunga . <i>Facies lon-|ga .</i> Dughi obraş.
 Faccia pallida , e mife-|rabile . <i>Facies
 miserabilis .</i> Lize bljedo , i◊tuşcno .
 Proverb. del volgo . <x>Takògàje liza ,◊kô da ni
 na Boşcich njie slana işio .</x>◊Defcr
 .◊<x>Şagugnastio , i obrasto◊Mustafàse vas ukaşa
 ;◊Lize suho , bljedo , i tmasto ,◊Pun skoncjagna
 , pun poraşa .</x> Ofm.◊Defcr .◊<x>Şagugnastio ,
 i şarasto◊Ovi u kosah vasse vidi ;◊Postarano lize
 , i tmasto ,◊Kaşce od koşce suhor blidi .</x>
 Ofm. Faccia◊fpaventevole . <i>Facies horrenda
 .</i> Obraş ftra-|fcjan . Faccia fcolorita .
 <i>Facies decolora-|ta .</i> Lize potamgneno .
 Faccia da galan-|tuomo . <i>Facies liberalis
 .</i> Lize plemenito .◊Faccia piena di brofole .
 <i>Puftulata rubro fa-|cies .</i> Obraş
 puhvicjaft . Obraş zargljenîm◊puhvizam pobjen.
 Faccia brutta . <i>Facies◊deformis .</i> Gard
 obraş. Grubo lize . Nèsklad-|no oblicje
 .◊<x>Pogrescpana , stara liza ,◊Sbabgljen Zarnaz
 , gard savisce .</x> Ofm. A◊prima faccia .
 <i>Prima fpecie .</i> Na parvi po-|gled . A
 faccia a faccia . <i>Facie ad faciem .</i>◊Lizem
 k' lizu . Lize na lizu . Non ha fac-|cia di
 parlare . <i>Erubefcit loqui .</i> Nèima o-|bras

, şa govoriti . Sràngàje , govoriti . Faccia◊con
 rughe . <i>Facies rugofa .</i> Lize pogrefcpano
 .◊<x>Pogrescpana , stara liza ,◊Sbabgljen Zarnaz
 , gard savisce .</x> Ofm.◊Guardar uno in faccia .
 <i>Alicujus os intue-|ri .</i> Gledati tkoga u
 obraş. Gledat tkoga◊u occhi . Şapişcdrîtfe u
 cigovo lize , alli u tko-|ga . Gli affari avevano
 mutata faccia . <i>Ma-|gna erat rerum facta
 commutatio .</i> Velikà-|je bila promjena u
 ftvarjeh , alli u pos-|ljeh . Bilèfu ftvari
 promjenile sliku . Posli-|sùfe bili priobraşili .
 Uomo fenza faccia ,◊vale fenza vergogona .
 <i>Impudens .</i> Govjek◊bèşobraşàn , alliti
 beşocjan.

Faldáto , che ha la falda . <i>Fimbriatus , ta ,◊tum
 .</i> Skutaft , skutafta , skutafto .

Fanciullezza . V. Puerilità .

Faticofiffimo . <i>Laboriofiffimus , ma , mum .</i>
 Pri-|trudni , dna , dno . Najtrudnji , a , e .
 Naj-|mucnji , jia , jie .

Fatto , add. <i>Factus , ta , tum .</i> Ucignen , ena
 ,◊eno . Proverb. <x>Velèje od recenoga do ucigne-
 |noga .</x> Dal detto al fatto è un gran trat-|to
 . Tvoren. Satvoren , na , no . Dofpi-|ven , ena ,
 eno . Svarfcen , na , no . Do-|po il fatto .
 <i>Re facta .</i> Po našada . Po ftvo-|ru . Po
 djellu .◊<x>Ersèje pak' şaman po stvoru kajati
 .</x> Ragn.◊Gljub. Mezzo fatto . <i>Semifactus ,
 ta , tum .</i>◊Nàpo ucignen. Nàpola ucignen.

Polùci-|gnen , ena , eno . Non fatto .
<i>Infectus , òcta , ctum .</i> Nèucignen , ena ,
 eno . Nèftvo-|ren , ena , no . Nèfvarfcen , na ,
 no . Nè-|dofpiven , ena , eno . Fatto
 maeftrevolmen-|te . *<i>Fabrefactus , cta , ctum .</i>*
 Skladnòucignen . òHitroucignen , alli
 iſhitren , na , no . Fat-|to alla groffa .
<i>Pinguì minerva factus .</i> O-|tefan na debelo
 . Ucignen na debelo . Fat-|to a lumaca . *<i>In
 cochleam retortus .</i>* Uci-|gnen na fpuſc , na
 iſvarte , u ſavitak . Fat-|to , cioè ftagionato ,
 e maturo . *<i>Maturus , òra , rum .</i>* Şrio ,
 rela , elo . Mezzo ftagio-|nato , mezzo fatto .
<i>Semicoctus , ta , tum .</i> òPolùşrio , rela ,
 lo . Nàpo şrio . Fatto a fca-|la . *<i>Gradatus ,
 ta , tum .</i>* Ucignen na liſtve .
 Fifchiamento , il fifchiare . *<i>Sibili editiæ .</i>*
 Şviſc-|dagne , gna . n. Şàşviſcdagne , gna .
 n. òŞviſcd , da . m.
 Fondatiffimo , molto bene fondato . *<i>Funda-
 |tiffimus , a , um .</i>* Prem utemeglñi , na , no
 .
 Fortúna , borrafca di tempo . V. Tempeſta . òFortuna ,
 cioè forte . *<i>Fortuna , næ . f.</i>* Sre-|chja ,
 chje . f. ò<x>Tako srechja svîm puſiva òU promjeni
 vjecnjeh dobaa òPromjenjivat svegh uſciva òRoba u
 Kraglja , Kraglja u roba .</x> Ofm. òFarfi la
 fortuna . *<i>Felicitatem fibi parare .</i>*
 òGraditi febi frechju . Sgotoviti frechju .

Ca-|dere in baffa fortuna . <i>Labafco , fcis
 .</i> Do-|chi na magne . Pařtit u potrebu , &c.
 V.◊Impoverire . Metter una cofa alla fortu-|na .
 <i>Aliquid fortunæ committere .</i> Spuřtiti
 na◊frechju . Fortuna felice . V. Prosperità
 .◊Fortuna cattiva . V. Avverfità . Per
 mala◊fortuna . <i>Fato infelici .</i> Po hudoj
 frechi .◊Per fortuna . V. Cafualmente .

Frecciäre , e faettare , tırrar freccie , o ferir◊con
 freccie . <i>Sagittas emittere .</i> Strjegljati
 ,◊gljam , glja◊fam . Strjeliti , lim , li◊fam
 .◊Strieglivati .◊<x>Jedan poghled bludno
 strigljen ,◊Sladak posmjeh , rjecza od meda ,
 &c.</x>◊Rařm. Strjelivati ftrjelom . Uftrjeliti .

Frotta , moltitudine , catterva . <i>Catterva , væ
 ,◊f.</i> Jatto , tta . m. Mnořc , ořci . f.
 Ghljud-|ftvo , ftva . m. Skup , pa . m. Cetta ,
 tte .◊f. V. Moltitudine . In frotta , avv.
 <i>Confer◊tim .</i> U jattu . Skuppo . U jedno .
 U cetti .

11.2. Izdvojene natuknice u izvornom obliku

U izvornom prikazu teksta, u tekstualnom dijelu tekst označen HTML oznakama `<i> . . . </i>` označen je kosim, crnim (automatski označenim) tekstom, a tekst označen HTML oznakama `. . .` koji je izvorno pisan kosim crnim slovima označen je kosim, tamnoplavim slovima (dark blue - text2) kako se ne bi izgubila razlikovnost oznaka „i“ odnosno „x“.

Znak za latinsko dugo s f u kurentnom kosom obliku *f* u della Bellinom izvornom tekstu upotrijebljen je za označavanje slova z u hrvatskim riječima (u kombinaciji s nekim drugima znači i š i ž). Taj jeznak u tekstu za bazu pisan znakom ş, a u ovom je prikazu zamijenjen znakom sličnim izvorniku: *f* (prema Unicodu latinsko malo slovo Esh - kôd 643). Kurentni oblik toga znaka je *f* (koso *f*) koji je grafički u kosoj verziji jednak dugom latinskom slovu s (*f*), a verzalni Σ. U izvorniku della Bellinog rječnika ne razlikuju se veliko i malo slovo *f*, nego je uvijek jednako, a najbližije je izvorniku u kosoj (italic) verziji *f*. Zato je i upotrijebljeno slovo ş (kurent) i *f* (verzal) da bi se u eventualnim kasnijim analizama vidjela ta razlika. Zato smo, da bi u ovom prikazu bilo najbližije izvorniku, a da se u digitalnom tekstu zadrži razlikovnost, veliko slovo Ş zamijenili kosim slovom *f* koje na dnu ima viticu.

U ovom prikazu prijelaz na novu stranicu označen je neprekinutom linijom „—————“, a prijelaz na novi stupac crtkanom linijom „- - - - -“.

| Izvorni tekst rječnika | Izvorni slog (slika) rječnika |
|--|--|
| Abbarbicare , per metter radici . V. Barbare . Radicare . | Abbarbicare , per metter radici . V. Barbare . Radicare . |
| Abbondantissimamente . avv. <i>Abundantissime</i> . Priobilato . Priobilno . Pribogato . | Abbondantissimamente . avv. <i>Abundantissime</i> . Priobilato . Priobilno . Pribogato . |
| Acchetatissimo . <i>Sedatissimus</i> , a . m. Priutascen , na , no . Primiran , irna , irno . Najuta/cenji , a , e . Najsmirenji , a , e . | Acchetatissimo . <i>Sedatissimus</i> , a . m. Priutascen , na , no . Primiran , irna , irno . Najuta/cenji , a , e . Najsmirenji , a , e . |
| Acetire , ed inacetire , farli aceto . <i>Acesco</i> , <i>scis</i> . U/oftitise , oftimsfe , tiòsàmse . U/cechise , u/ci/cèmse u/cegaòsàmse . <i>scighnuti</i> fe , ghgnivàmse , <i>scighnuòsàmse</i> . U/gljuti- tise , gljutimsfe , gljutiòsàmse . Profuknuti , gnivam , profuknuòsam . Pochi na ozat . | Acetire , ed inacetire , farli aceto . <i>Acesco</i> , <i>scis</i> . U/oftitise , oftimsfe , tiòsàmse . U/cechise , u/ci/cèmse u/cegaòsàmse . <i>scighnuti</i> fe , ghgnivàmse , <i>scighnuòsàmse</i> . U/gljuti- tise , gljutimsfe , gljutiòsàmse . Profuknuti , gnivam , profuknuòsam . Pochi na ozat . |
| Pochi na <i>scizzu</i> , li dice del vino . Acetire del pane . U/kvasitise , livàmse , u/kva- tiòsàmse . Prikisnuti , lgnivam , prikisnuò- sam . Acetirli del latte . U/vifctàtse , fctu- jèmse , u/vifctàòsàmse . Profuknuti , gni- vam , profuknuòsam . v. g. Mljekòje pro- fuknulo . U/kisjelitise , livàmse , u/kifelilòsàm- se . Far acetire . <i>In acetum convertere</i> . Ci- niti u/oftiti . Ciniti , dàse u/ofti . U/kvasiti kruh ; far acetir il pane . Ukisjeliti , far acetir il latte . L' acetire , il farli aceto . <i>In acetum mutatio</i> . U/ostegne , egna . n. Il farli aceto del pane . U/kvafegne , gna . n. | Pochi na <i>scizzu</i> , li dice del vino . Acetire del pane . U/kvasitise , livàmse , u/kva- tiòsàmse . Prikisnuti , lgnivam , prikisnuò- sam . Acetirli del latte . U/vifctàtse , fctu- jèmse , u/vifctàòsàmse . Profuknuti , gni- vam , profuknuòsam . v. g. Mljekòje pro- fuknulo . U/kisjelitise , livàmse , u/kifelilòsàm- se . Far acetire . <i>In acetum convertere</i> . Ci- niti u/oftiti . Ciniti , dàse u/ofti . U/kvasiti kruh ; far acetir il pane . Ukisjeliti , far acetir il latte . L' acetire , il farli aceto . <i>In acetum mutatio</i> . U/ostegne , egna . n. Il farli aceto del pane . U/kvafegne , gna . n. |
| Adiratamente . V. Iratamente . | Adiratamente . V. Iratamente . |
| Affreddare . V. Raffreddare . | Affreddare . V. Raffreddare . |
| Affrica , ed Africa , una delle quattro parti del Mondo . <i>Africa</i> , <i>cæ</i> . f. Afrika , ke . f. Afrika , i Poludne vighjascèse , ne bes sufaa priklopjeno orufcjem , i poluvjerstvom od Vandalaa . Gjur. S. Ben. | Affrica , ed Africa , una delle quattro parti del Mondo . <i>Africa</i> , <i>cæ</i> . f. Afrika , ke . f. Afrika , i Poludne vighjascèse , ne bes sufaa priklopjeno orufcjem , i poluvjerstvom od Vandalaa . Gjur. S. Ben. |
| Agguagliamento , cioè l' agguagliare . <i>Æqua-</i> | Agguagliamento , cioè l' agguagliare . <i>Æqua-</i> |
| <i>tio</i> , <i>nis</i> . f. Jednacegne , gna . n. Takme- gne , gna . n. Istakmegne , gna . n. Ifjed- nacegne , gna . n. | <i>tio</i> , <i>nis</i> . f. Jednacegne , gna . n. Takme- gne , gna . n. Istakmegne , gna . n. Ifjed- nacegne , gna . n. |
| Alato . add. che ha ale . <i>Alatus</i> , a , um . Krilat , ta , to . Krilaft , fta , fto . Krilataft , fta , fto . Letufcti , fcta , fcto . Letufcte , fcta . n. fust . I gnom' platta svud letuscta Po planini varti , i vjie . Gjur. Raf. Chi ha il piede alato , o veloce . <i>Alipes</i> , dis . Krilònogh , noga , nogo . | Alato , add. che ha ale . <i>Alatus</i> , a , um . Krilat , ta , to . Krilaft , fta , fto . Krilataft , fta , fto . Letufcti , fcta , fcto . Letufcte , fcta . n. fust . I gnom' platta svud letuscta Po planini varti , i vjie . Gjur. Raf. Chi ha il piede alato , o veloce . <i>Alipes</i> , dis . Krilònogh , noga , nogo . |
| Allegatore , chi allega , chi cita . <i>Qui afferi</i> . Prika/cjujuchi , uchja , uchje . Prika/alaz , aoza . m. Ifvoditegl , glja . n. | Allegatore , chi allega , chi cita . <i>Qui afferi</i> . Prika/cjujuchi , uchja , uchje . Prika/alaz , aoza . m. Ifvoditegl , glja . n. |

| | |
|---|--|
| <p>Allegro , suft. cioè l' uomo di natura al-
legro . <i>Vir ingenio hilari</i> . Vefegljak , aka . m.
<i>Gljubàvje , fciivjeti s' druſcinom , junazi ,</i>
<i>Rujno vinze piti s' dobrjem vefegliazi ,</i>
Tir. Vefe le naravi . Vefe le chjudi .</p> |  <p>Allegro , suft. cioè l' uomo di natura al-
legro . <i>Vir ingenio hilari</i> . Vefegljak , aka . m.
<i>Gljubàvje , fciivjeti s' druſcinom , junazi ,</i>
<i>Rujno vinze piti s' dobrjem vefegliazi ,</i>
Tir. Vefe le naravi . Vefe le chjudi .</p> |
| <p>Allentare , far lento . <i>Laxo , xas</i> . Popuſctati ,
tavam , tiòſam . Popuſtiti , vam , ſtiòſam .
Oslabiti , vam , biòſam . Opruſciti , ſci-
vam , ſciòſam . Odpuſtiti , vam , ſtiòſam .
v. g. Odpuſctavam paas . Allento la cinta .
Allentare troppo . <i>Nimis relaxare</i> . Raſla-
biti , slabivam , biòſam . Raſpuſtiti , puſti-
vam , tiòſam . Raſpruſciti , ſcivam , ſciò-
ſam . Allentare poco . <i>Parum laxare</i> . Po-
puſtiti , puſctavam , tiòſam . Popruſciti ,
ſcjujem , ſciòſam .</p> |  <p>Allentare , far lento . <i>Laxo , xas</i> . Popuſctati ,
tavam , tiòſam . Popuſtiti , vam , ſtiòſam .
Oslabiti , vam , biòſam . Opruſciti , ſci-
vam , ſciòſam . Odpuſtiti , vam , ſtiòſam .
v. g. Odpuſctavam paas . Allento la cinta .
Allentare troppo . <i>Nimis relaxare</i> . Raſla-
biti , slabivam , biòſam . Raſpuſtiti , puſti-
vam , tiòſam . Raſpruſciti , ſcivam , ſciò-
ſam . Allentare poco . <i>Parum laxare</i> . Po-
puſtiti , puſctavam , tiòſam . Popruſciti ,
ſcjujem , ſciòſam .</p> |
| <p>Alternare , cioè far una coſa a vicenda . <i>Al-
terno , as , avi , atum</i> . Iſmjenitiſe , gnu-
jèmſe , niòſàmſe . Promjenitiſe , gnujèm-
ſe , niòſàmſe . Ciniti ſctògod na iſmjenu .
v. g. Pripovjedati na iſmjenu .
<i>Na iſmjenu kí pritvarasc</i>
<i>Noch u ſvjetlos dan u tmine</i> , Gjor. Raſ.
Iſmjenitiſe u pripovjedagnu . Freq. Iſmje-
nivatiſe . Promjegnivatiſe .</p> |  <p>Alternare , cioè far una coſa a vicenda . <i>Al-
terno , as , avi , atum</i> . Iſmjenitiſe , gnu-
jèmſe , niòſàmſe . Promjenitiſe , gnujèm-
ſe , niòſàmſe . Ciniti ſctògod na iſmjenu .
v. g. Pripovjedati na iſmjenu .
<i>Na iſmjenu kí pritvarasc</i>
<i>Noch u ſvjetlos dan u tmine</i> , Gjor. Raſ.
Iſmjenitiſe u pripovjedagnu . Freq. Iſmje-
nivatiſe . Promjegnivatiſe .</p> |
| <p>Alzare , inalzare . <i>Effero , ers</i> . Uſdighnuti ,
diſcem , dighnuòſam . Uſpeti , uſpignem ,
uſpeòſam .
<i>Glasoviti nu pisaozi</i>
<i>Uſpeſcèjih na toliko</i> , Ofm. Podighnuti ,
iſcem , dighnuòſam . Uſviſiti , ſcjujem ,
ſiòſam .
<i>ſcena ohola ne misli ino</i> ,
<i>Nego uſviſit ſebe , i ſina</i> . Ofm. Pripeti ,
pignam , peòſam .
<i>Kí Mustafu opet ſcele</i>
<i>Na ſto zarski ſvoj pripeti</i> . Ofm. Pro-
peti , nem , peòſam .</p> |  <p>Alzare , inalzare . <i>Effero , ers</i> . Uſdighnuti ,
diſcem , dighnuòſam . Uſpeti , uſpignem ,
uſpeòſam .
<i>Glasoviti nu pisaozi</i>
<i>Uſpeſcèjih na toliko</i> , Ofm. Podighnuti ,
iſcem , dighnuòſam . Uſviſiti , ſcjujem ,
ſiòſam .
<i>ſcena ohola ne misli ino</i> ,
<i>Nego uſviſit ſebe , i ſina</i> . Ofm. Pripeti ,
pignam , peòſam .
<i>Kí Mustafu opet ſcele</i>
<i>Na ſto zarski ſvoj pripeti</i> . Ofm. Pro-
peti , nem , peòſam .</p> |

fàsve da od gne , kad oholu
 Glavu istakmit k' Viscgnem' , prope ,
 Starmoglàvse buscj dolu
 Vjecnjeh ogagn' u potope . Ofm. Propi-
 gnati , gnem , pignòlam . Pegnati , gnem ,
 gnaòlam
 Cjem slòsardi smaj na mene glavu pegne ,
 Gjor. Pfal. Podignuti , di/cem , dignuòlam .
 Dignuti , di/cem , nuòlam . Vifiti , višim ,
 vifiòlam
 Nakitjeno dubjem celo
 Put nebesaa vise , i steru . Palm. Chr.
 V. Inalzare . Alzar in alto . In subline
 erigere . U/dignuti , nivam , nuòlam . Dig-
 nuti , di/cem , nuòlam
 Ter dignite , ter upràvte
 Vascju pamet put nebesaa . Palm. Chr.
 Alzare le mani al cielo . Tollere manus ad
 cælum . U/dignuti ruke , k' nebu . Uprošiti
 rukée k' nebu
 Tim put dvora nebeskoga
 Occi uprje , rukée uprostj . Palm. Chr.
 K' nebu pru/citi rukée , pru/ciam , sciòlam .
 Alzare di sotto . Elevare . Podignuti . Po-
 diti , illi dignuti . Podpodignuti , di/cem ,
 gaòlam . Alzar qualcheduno . Erigere ali-
 quem . U/dignuti , alliti podignuti tkoga .
 Alzare la mano per percuotere . Elevare
 manum ad percutiendum samahnuti , mah-
 nujem , nuòlam . Freq. samahivati , hivam ,
 vòlam . Alzare da terra qualche cosa .
 Elevare . Dokuciti s' tle kojùgod stvar ,
 cjujem , ciòlam . Freq. Dokucivati , civam ,
 viòlam . Alzarfi in punta di piedi . Consi-
 flere digitis . Propetise na parste , propi-
 gnèmse , peòsàmse . Ispetise , pignàmse ,
 peòsàmse . Freq. Ispignatise , gnàmse , gnaò-
 sàmse . Pripetise na parste , pripignèmse ,
 pripèòsàmse . Alzarfi da letto . Surgere e
 lecto . Uftati , uftajem , staòlam . Dignutise ,
 alliti uftajati i/ odra . Alzarfi in piedi . Af-
 surgo , gis . Dignutise , di/cèmse , nuòsà-
 mse . Uftati na noghee . Podvignutise , vi-
 scèmse , nuòsàmse . Uftati , tajem , taòlam .
 A Dilaver dogovorj ,
 Parvi Vefier parvi usta . Ofm. Alzarfi .
 Surgere . U/dignutise , di/cèmse , gnuòsà-
 mse . Uftati , ajem , aòlam . Dignutise , di-
 scèmse , dvignuòsàmse . Podignutise , scè-
 mse , nuòsàmse
 fa tvojùchje vechju szjennu
 Zarkve podvignut neifbrojene . Pal. Chr.
 Ufraštjeti , illi ufraštiti
 Do zarskoga isctom stola ,
 Odkli pade , da ufrašti . Ofm. Chi alza .

Erigens , tis . U/dignitegl , glja . m. Podi-
 gnitegl , glja . m. Podi/cjuchi , chja , chje ;
 e il fem. Podignitegliza , ze . f. U/di-
 gnitegliza , ze . f.

fàsve da od gne , kad oholu
 Glavu istakmit k' Viscgnem' , prope ,
 Starmoglàvse buscj dolu
 Vjecnjeh ogagn' u potope . Ofm. Propi-
 gnati , gnem , pignòlam . Pegnati , gnem ,
 gnaòlam
 Cjem slòsardi smaj na mene glavu pegne ,
 Gjor. Pfal. Podignuti , di/cem , dignuòlam .
 Dignuti , di/cem , nuòlam . Vifiti , višim ,
 vifiòlam
 Nakitjeno dubjem celo
 Put nebesaa vise , i steru . Palm. Chr.
 V. Inalzare . Alzar in alto . In subline
 erigere . U/dignuti , nivam , nuòlam . Dig-
 nuti , di/cem , nuòlam
 Ter dignite , ter upràvte
 Vascju pamet put nebesaa . Palm. Chr.
 Alzare le mani al cielo . Tollere manus ad
 cælum . U/dignuti rukée k' nebu . Uprošiti
 rukée k' nebu
 Tim put dvora nebeskoga
 Occi uprje , rukée uprostj . Palm. Chr.
 K' nebu pru/citi rukée , pru/ciam , sciòlam .
 Alzare di sotto . Elevare . Podignuti . Po-
 diti , illi dignuti . Podpodignuti , di/cem ,
 gaòlam . Alzar qualcheduno . Erigere ali-
 quem . U/dignuti , alliti podignuti tkoga .
 Alzare la mano per percuotere . Elevare
 manum ad percutiendum samahnuti , mah-
 nujem , nuòlam . Freq. samahivati , hivam ,
 vòlam . Alzare da terra qualche cosa .
 Elevare . Dokuciti s' tle kojùgod stvar ,
 cjujem , ciòlam . Freq. Dokucivati , civam ,
 viòlam . Alzarfi in punta di piedi . Consi-
 flere digitis . Propetise na parste , propi-
 gnèmse , peòsàmse . Ispetise , pignàmse ,
 peòsàmse . Freq. Ispignatise , gnàmse , gnaò-
 sàmse . Pripetise na parste , pripignèmse ,
 pripèòsàmse . Alzarfi da letto . Surgere e
 lecto . Uftati , uftajem , staòlam . Dignutise ,
 alliti uftajati i/ odra . Alzarfi in piedi . Af-
 surgo , gis . Dignutise , di/cèmse , nuòsà-
 mse . Uftati na noghee . Podvignutise , vi-
 scèmse , nuòsàmse . Uftati , tajem , taòlam .
 A Dilaver dogovorj ,
 Parvi Vefier parvi usta . Ofm. Alzarfi .
 Surgere . U/dignutise , di/cèmse , gnuòsà-
 mse . Uftati , ajem , aòlam . Dignutise , di-
 scèmse , dvignuòsàmse . Podignutise , scè-
 mse , nuòsàmse
 fa tvojùchje vechju szjennu
 Zarkve podvignut neifbrojene . Pal. Chr.
 Ufraštjeti , illi ufraštiti
 Do zarskoga isctom stola ,
 Odkli pade , da ufrašti . Ofm. Chi alza .

Erigens , tis . U/dignitegl , glja . m. Podi-
 gnitegl , glja . m. Podi/cjuchi , chja , chje ;
 e il fem. Podignitegliza , ze . f. U/di-
 gnitegliza , ze . f.

| | |
|---|---|
| <p>Ambasciata , quel che riferisce l' ambasciadore . <i>Mandatum</i> , ti . n. Poruka , ke . f. Far l' ambasciata . <i>Mandata exponere</i> . Prikasati poruku . Opovidjeti poruku . Portare l' ambasciata . <i>Mandata ad aliquem afferre</i> . Donjeti tkomu poruku . Mandar ambasciata ad uno per un altro . <i>Mandata alicui ad aliquem dare</i> . Poruciti sctògod tkomu po tkomùgod .</p> | <p>Ambasciata , quel che riferisce l' ambasciadore . <i>Mandatum</i> , ti . n. Poruka , ke . f. Far l' ambasciata . <i>Mandata exponere</i> . Prikasati poruku . Opovidjeti poruku . Portare l' ambasciata . <i>Mandata ad aliquem afferre</i> . Donjeti tkomu poruku . Mandar ambasciata ad uno per un altro . <i>Mandata alicui ad aliquem dare</i> . Poruciti sctògod tkomu po tkomùgod .</p> |
| <p>Amello , erba . <i>Amellus</i> . Divglisciza , ize . f. Rutta divja , vje . f.</p> | <p>Amello , erba . <i>Amellus</i> . Divglisciza , ize . f. Rutta divja , vje . f.</p> |
| <p>Ammonimento , ammonizione . <i>Admonitio</i> , nis . f. Nauk , ka . m. Svjetovagne , gna . n. Uvjecbagne , gna . n. Nagovaragne , gna . n. Napomenutje , tja . n. Osvjestjegne , gna . n. Posvjestjegne , gna . n. Svjestjegne , gna . n.</p> | <p>Ammonimento , ammonizione . <i>Admonitio</i> , nis . f. Nauk , ka . m. Svjetovagne , gna . n. Uvjecbagne , gna . n. Nagovaragne , gna . n. Napomenutje , tja . n. Osvjestjegne , gna . n. Posvjestjegne , gna . n. Svjestjegne , gna . n.</p> |
| <p>Anelletti , a' quali si affibbiano gli uncinelli delle vesti . <i>Anuli vestiarii</i> . Spone , naa . f. plur.</p> | <p>Anelletti , a' quali si affibbiano gli uncinelli delle vesti . <i>Anuli vestiarii</i> . Spone , naa . f. plur.</p> |
| <p>Anello , che si porta in dito . <i>Anulus</i> , li . m. Parften , na . m. Anello senza pietra . <i>Anulus purus</i> . Parften bes' ofcza . Anello spofalizio . <i>Anulus pronubus</i> . Rukovni parften . Parften blagofovjeni . farucni parften . scenidbeni parften . Parften pirni . <i>I prid bojnóm sabgljom gljutóm Parsten pirni cjuše uljesti</i> . Ofm. Anello d' oro . <i>Anulus aureus</i> . flatni parften . Luogo , o parte dell' anello , in cui s' incastra la gemma . <i>Pala</i> , æ . f. Bok od parstena . I/nofito od parstena . Anello ingemmato , che ha la gemma . <i>Anulus gemmatus</i> . Parften s' ofczem , s' draghjem kamenom . Anello da sigillare . <i>Anulus signatorius</i> . Pecjatni parften . Parften pecjatnik . Anello da cucire . V. Ditale . Anello , o anelletto da cortina . <i>Anuli cortinales</i> . Biocjuh , cjuha . m. Anello da catena . <i>Catenæ anulus</i> . Biocjuh , ha . m. Porfi , o metterfi l' anello . <i>Anulum induere</i> . Staviti parften na ruku . Parften na</p> | <p>Anello , che si porta in dito . <i>Anulus</i> , li . m. Parften , na . m. Anello senza pietra . <i>Anulus purus</i> . Parften bes' ofcza . Anello spofalizio . <i>Anulus pronubus</i> . Rukovni parften . Parften blagofovjeni . farucni parften . scenidbeni parften . Parften pirni . <i>I prid bojnóm sabgljom gljutóm Parsten pirni cjuše uljesti</i> . Ofm. Anello d' oro . <i>Anulus aureus</i> . flatni parften . Luogo , o parte dell' anello , in cui s' incastra la gemma . <i>Pala</i> , æ . f. Bok od parstena . I/nofito od parstena . Anello ingemmato , che ha la gemma . <i>Anulus gemmatus</i> . Parften s' ofczem , s' draghjem kamenom . Anello da sigillare . <i>Anulus signatorius</i> . Pecjatni parften . Parften pecjatnik . Anello da cucire . V. Ditale . Anello , o anelletto da cortina . <i>Anuli cortinales</i> . Biocjuh , cjuha . m. Anello da catena . <i>Catenæ anulus</i> . Biocjuh , ha . m. Porfi , o metterfi l' anello . <i>Anulum induere</i> . Staviti parften na ruku . Parsten na</p> |
| <p>taknuti , ticem , nuòfam , illi natakòfam . Levarfi l' anello . <i>Deponere anulum</i> . Skinuti , kinivam , nuòfam . Staknuti , knivam , knuòfam . Smaknuti , knivam , knuòfam parften s' ruke . Portar l' anello in dito . <i>Digito anulum gestare</i> . Nofiti , nofim , nofiòfam parften na ruzi . Fare , o lavorare anelli . <i>Fabricare anulos</i> . Kovati , kujem , kovòfam parstene .</p> | <p>taknuti , ticem , nuòfam , illi natakòfam . Levarfi l' anello . <i>Deponere anulum</i> . Skinuti , kinivam , nuòfam . Staknuti , knivam , knuòfam . Smaknuti , knivam , knuòfam parften s' ruke . Portar l' anello in dito . <i>Digito anulum gestare</i> . Nofiti , nofim , nofiòfam parften na ruzi . Fare , o lavorare anelli . <i>Fabricare anulos</i> . Kovati , kujem , kovòfam parstene .</p> |
| <p>Anguillaia , luogo , o ridotto di anguille . <i>Anguillarum receptaculum</i> . Jeguglifcte , fcta . n. Mjesto jegugljam' plodno .</p> | <p>Anguillaia , luogo , o ridotto di anguille . <i>Anguillarum receptaculum</i> . Jeguglifcte , fcta . n. Mjesto jegugljam' plodno .</p> |

| | |
|---|---|
| <p>Annebbiamento , lo annebbarfi . <i>Nebula te-
gi</i> .Omaghlegne , gna . n. Namaghlegne ,
gna . n.</p> | <p>Annebbiamento , lo annebbarfi . <i>Nebula te-
gi</i> . Omaghlegne , gna . n. Namaghlegne ,
gna . n.</p> |
| <p>Anno . <i>Annus</i> , <i>ni</i> . <i>m</i> . Godifcte , fcta . n.
Godina , ne . f.
<i>Evo gre pet godin' , dàsam gne sluga ja ;</i>
Scifc. Ljetto , ta . n. Anno nuovo . <i>Annus
incipiens</i> . Mlado ljetto . Novo godifcte .
Un' anno . <i>Annus</i> , <i>i</i> . <i>m</i> . Godifcte danaa .
Jedno godifcte . Mezzo anno . <i>Semiannus</i> ,
<i>ni</i> . <i>m</i> . Po godifcta . indecl . Po ljetta . Po-
loviza godifcta . Un' anno , e mezzo . <i>Sef-
quiannus</i> , <i>ni</i> . <i>m</i> . Godifcte i po . Ljetto i
po . D' anno in anno . <i>In annos singulos</i> .
Od godifcta do godifcta . Od ljetta do ljet-
ta . D' un' anno . <i>Annulus</i> , <i>a</i> , <i>um</i> . Od
jednoga godifcta . Od jednoga ljetta . Jed-
nòljetni , na , no . Godifctni , na , no ,
illi od godifcta . Jednògodni , na , no .
Jednògodaz , dza . <i>m</i> . Di due anni . <i>Bi-
mus</i> , <i>a</i> , <i>um</i> , Dvòljetni , na , no . Dvò-
godni , na , no . Dvògodaz , dza . <i>m</i> . Od
dva godifctaa . Di tre anni . <i>Trimus</i> , <i>a</i> ,
<i>um</i> . Tròljetni , tna , tno . Trechjak , aka .
<i>m</i> . Di quattro anni . <i>Quadrinus</i> , <i>a</i> , <i>um</i> .
Cetveròljetni , na , no . Cetverògodni ,
Cetverògodaz , dza . <i>m</i> . Cetvartak , aka .
<i>m</i> . fust. Di cinque anni . <i>Quinquennis</i> . &
<i>hoc quinquenne</i> . Peteròljetni , na , no .
Peterògodni , na , no . Peterògodaz , dza .
Petak , aka . <i>m</i> . fust. Di sei anni . <i>Se-
xennis</i> , <i>sexenne</i> . Scefteròljetni , na , no .
Scefterògodni , na , no . Scefterògodaz ,
dza . <i>m</i> . Sceftak , aka . <i>m</i> . fust. Di sette
anni . <i>Septennis</i> . & <i>hoc ne</i> . Sedmeròljet-
ni , tna , tno . Sedmerògodni &c. Sedmak ,
aka . <i>m</i> . Di otto anni . <i>Octennis</i> . & <i>hoc
ne</i> . Ofmeròljetni , tna , tno . Ofmerògodni
&c. Ofmak , aka . <i>m</i> . fust. Di nove anni .
<i>Novennis</i> , & <i>hoc ne</i> . Deveteròljetni , na ,
no . Deveterògodni &c. Devetak , aka .
<i>m</i> . Deveterògodaz , dza . <i>m</i> . Di dieci an-
ni . <i>Decennis</i> , & <i>hoc ne</i> . Defeteròljetni ,
na , no . Defeterògodni , na , no . Defe-
tak , aka . <i>m</i> . fust. Due anni . <i>Biennium</i> ,
<i>ii</i> . <i>n</i> . Dva godifcta . Dvie godine danaa .
Dva ljetta . Tre anni , tempo di tre an-
ni . <i>Triennium</i> , <i>nii</i> . <i>n</i> . Tri godifcta . Tri
godine danaa . Tri ljetta . Tre anni com-
piti , o finiti . <i>Trieteris plena</i> . Tri punna
godifcta . Tri zjela godifcta . Quattro an-</p> | <p>Anno . <i>Annus</i> , <i>ni</i> . <i>m</i> . Godifcte , fcta . n.
Godina , ne . f.
<i>Evo gre pet godin' , dàsam gne sluga ja ;</i>
Scifc. Ljetto , ta . n. Anno nuovo . <i>Annus
incipiens</i> . Mlado ljetto . Novo godifcte .
Un' anno . <i>Annus</i> , <i>i</i> . <i>m</i> . Godifcte danaa .
Jedno godifcte . Mezzo anno . <i>Semiannus</i> ,
<i>ni</i> . <i>m</i> . Po godifcta . indecl . Po ljetta . Po-
loviza godifcta . Un' anno , e mezzo . <i>Sef-
quiannus</i> , <i>ni</i> . <i>m</i> . Godifcte i po . Ljetto i
po . D' anno in anno . <i>In annos singulos</i> .
Od godifcta do godifcta . Od ljetta do ljet-
ta . D' un' anno . <i>Annulus</i> , <i>a</i> , <i>um</i> . Od
jednoga godifcta . Od jednoga ljetta . Jed-
nòljetni , na , no . Godifctni , na , no ,
illi od godifcta . Jednògodni , na , no .
Jednògodaz , dza . <i>m</i> . Di due anni . <i>Bi-
mus</i> , <i>a</i> , <i>um</i> . Dvòljetni , na , no . Dvò-
godni , na , no . Dvògodaz , dza . <i>m</i> . Od
dva godifctaa . Di tre anni . <i>Trimus</i> , <i>a</i> ,
<i>um</i> . Tròljetni , tna , tno . Trechjak , aka .
<i>m</i> . Di quattro anni . <i>Quadrinus</i> , <i>a</i> , <i>um</i> .
Cetveròljetni , na , no . Cetverògodni ,
Cetverògodaz , dza . <i>m</i> . Cetvartak , aka .
<i>m</i> . fust. Di cinque anni . <i>Quinquennis</i> , &
<i>hoc quinquenne</i> . Peteròljetni , na , no .
Peterògodni , na , no . Peterògodaz , dza .
Petak , aka . <i>m</i> . fust. Di sei anni . <i>Se-
xennis</i> , <i>sexenne</i> . Scefteròljetni , na , no .
Scefterògodni , na , no . Scefterògodaz ,
dza . <i>m</i> . Sceftak , aka . <i>m</i> . fust. Di sette
anni . <i>Septennis</i> , & <i>hoc ne</i> . Sedmeròljet-
ni , tna , tno . Sedmerògodni &c. Sedmak ,
aka . <i>m</i> . Di otto anni . <i>Octennis</i> , & <i>hoc
ne</i> . Ofmeròljetni , tna , tno . Ofmerògodni
&c. Ofmak , aka . <i>m</i> . fust. Di nove anni .
<i>Novennis</i> , & <i>hoc ne</i> . Deveteròljetni , na ,
no . Deveterògodni &c. Devetak , aka .
<i>m</i> . Deveterògodaz , dza . <i>m</i> . Di dieci an-
ni . <i>Decennis</i> , & <i>hoc ne</i> . Defeteròljetni ,
na , no . Defeterògodni , na , no . Defe-
tak , aka . <i>m</i> . fust. Due anni . <i>Biennium</i> ,
<i>ii</i> . <i>n</i> . Dva godifcta . Dvie godine danaa .
Dva ljetta . Tre anni , tempo di tre an-
ni . <i>Triennium</i> , <i>nii</i> . <i>n</i> . Tri godifcta . Tri
godine danaa . Tri ljetta . Tre anni com-
piti , o finiti . <i>Trieteris plena</i> . Tri punna
godifcta . Tri zjela godifcta . Quattro an-</p> |

ni, lo spazio di quattro anni. *Quadriennium*, *nii*. *n.* Cetiri godifctaa. Cetiri godine danaa. Cinque anni. *Quinquennium*, *nii*. *n.* Pet godifctaa. Pet ljetaa. Pet godinaa.

Evo gre pet godin' . . . Scif. Che dura cinque anni. *Quinquennalis*, *le*. Petòljetni, na, no. Sei anni, lo spazio di sei anni. *Sexennium*, *nii*. *n.* Sces godifctaa. Sette anni. *Septennium*, *nii*. *n.* Sedam godifctaa. Sedmòljetni, na, no. Otto anni. *Octennium*, *nii*. *n.* Osam godifctaa. Osmòljetni, na, no. Nove anni. *Novennium*, *nii*. *n.* Devet godifctaa. Devetòljetni, na, no. Dieci anni. *Decennium*, *nii*. *n.* Deset godifctaa. Desetòljetni, na, no. Molti anni. *Multi anni*. Mnoga godifctaa. Velika godifctaa.

Ni mene rasdruscit' godiscta velika Nèmogu, da sluscit' ne buddem do vjika. Bur. Molti anni fa. *Multis ab hinc annis*. Mnogo godifctaa na sada. Jest mnogo godifctaa. Davno. Sono due anni. *Duobus ab hinc annis*. Jesu dua godifcta. Na sad dua godifcta. Di qui a dieci anni. *Ad decem annos*. Do deset godifctaa. Sa deset godifctaa. Nakon deset godifctaa. Ove deset godifctaa. Per un' anno. *In annum*. Sa godifcte. Per l' anno seguente. *In proximum annum*. Sa do ljetta. Sa do godifcta. Sa dohodno godifcte. Anno bifeftile, o del bifefto. *Annus intercalaris*. Pristupno godifcte. Pristupno ljetto. Anno Santo, ovvero, anno del Giubileo. *Annus saecularis*. Sveto godifcte. Anno lunare, quale dura un mese. *Annus lunaris*. Mjesecno godifcte. Anno solare. *Annus solaris*. Suncjano godifcte. Godifcte od dvanes mjesezaa. Di molti anni, colui; che ha molti anni. *Annosus*, *a*, *um*. Davnegni, na, no. Davgni, gna, gne.

Kjem sred viecne pechi davgne slobne duscee muce, i parfce. Palm. Chr. Pun godifctaa. Vremenit, ta, to.

Jurje musc vremenit, a pjesni pripjeva; Ekt. Raf. Una volta all' anno. *Semel in anno*. Jednòkrat na godifcte. Jedan put na godifcte. Jednom na godifcte. Finito l' anno. *Anno exacto*. Na ishòd godifcta. Na svarhu godifcta. Un' annò sì, e l' altro no. *Alternis annis*. Svako drugo godifcte. Svaka dva ljetta. Dentro al termine d' un' anno. *Intra annum*. U jedno godifcte. Più d' un' anno. *Ultra annum spatium*. Vifce od godifcta. Boglje od go-

ni, lo spazio di quattro anni. *Quadriennium*, *nii*. *n.* Cetiri godifctaa. Cetiri godine danaa. Cinque anni. *Quinquennium*, *nii*. *n.* Pet godifctaa. Pet ljetaa. Pet godinaa.

Evo gre pet godin' . . . Scif. Che dura cinque anni. *Quinquennalis*, *le*. Petòljetni, na, no. Sei anni, lo spazio di sei anni. *Sexennium*, *nii*. *n.* Sces godifctaa. Sette anni. *Septennium*, *nii*. *n.* Sedam godifctaa. Sedmòljetni, na, no. Otto anni. *Octennium*, *nii*. *n.* Osam godifctaa. Osmòljetni, na, no. Nove anni. *Novennium*, *nii*. *n.* Devet godifctaa. Devetòljetni, na, no. Dieci anni. *Decennium*, *nii*. *n.* Deset godifctaa. Desetòljetni, na, no. Molti anni. *Multi anni*. Mnoga godifctaa. Velika godifcta.

Ni mene rasdruscit' godiscta velika Nèmogu, da sluscit' ne buddem do vjika. Bur. Molti anni fa. *Multis ab hinc annis*. Mnogo godifctaa na sada. Jest mnogo godifctaa. Davno. Sono due anni. *Duobus ab hinc annis*. Jesu dua godifcta. Na sad dua godifcta. Di qui a dieci anni. *Ad decem annos*. Do deset godifctaa. Sa deset godifctaa. Nakon deset godifctaa. Ove deset godifctaa. Per un' anno. *In annum*. Sa godifcte. Per l' anno seguente. *In proximum annum*. Sa do ljetta. Sa do godifcta. Sa dohodno godifcte. Anno bifeftile, o del bifefto. *Annus intercalaris*. Pristupno godifcte. Pristupno ljetto. Anno Santo, ovvero, anno del Giubileo. *Annus saecularis*. Sveto godifcte. Anno lunare, quale dura un mese. *Annus lunaris*. Mjesecno godifcte. Anno solare. *Annus solaris*. Suncjano godifcte. Godifcte od dvanes mjesezaa. Di molti anni, colui; che ha molti anni. *Annosus*, *a*, *um*. Davnegni, na, no. Davgni, gna, gne.

Kjem sred viecne pechi davgne slobne duscee muce, i parfce. Palm. Chr. Pun godifctaa. Vremenit, ta, to.

Jurje musc vremenit, a pjesni pripjeva; Ekt. Raf. Una volta all' anno. *Semel in anno*. Jednòkrat na godifcte. Jedan put na godifcte. Jednom na godifcte. Finito l' anno. *Anno exacto*. Na ishòd godifcta. Na svarhu godifcta. Un' annò sì, e l' altro no. *Alternis annis*. Svako drugo godifcte. Svaka dva ljetta. Dentro al termine d' un' anno. *Intra annum*. U jedno godifcte. Più d' un' anno. *Ultra annum spatium*. Vifce od godifcta. Boglje od go-

| | |
|--|---|
| <p>disfeta . Vechje od godifceta . L' anno ap-
preſſo . <i>Post annum</i> . Drugo godifcte . Do
hodnega godifceta . Do godifceta . Do ljet-
ta . L' anno paſſato . <i>Anno præterito</i> . La-
ni . Lansko godifcte . Proſcjaſto godifcte ,
alli ljetto . L' anno venturo . <i>Anno futu-
ro</i> . Dohodno godifcte . Due anni ſono .
<i>Duobus ab hinc annis</i> . Onomlani . Jeſu
dvje godine danaa . Jeſu dva ljetta . Na
ſad dva godifctaa . Anno temperato . <i>An-
nus temperatus</i> . Meko godifcte . Anno in-
ſalubre . <i>Annus inſalubris</i> . Neſdravo go-
difcte . Anni della diſcrezione , ſono quel-
li , ne' quali uno diſcerne il bene dal ma-
le . <i>Sapientia anni</i> . Raſloſcna godifctaa .
Spametna godifctaa . Raſborna godifctaa .
Avere gl' anni della diſcrezione , ovvero
l' uſo della ragione . <i>Sapio , is , pui , itum</i> .
Imati raſborna godifctaa . Mochi raſabrati
ſlo od dobra . Biti ſtupio na ljetaa od
raſbora . Anni giovanili , o della gioventù
.
<i>Anni crescentes</i> . Mlada godifctaa .
Zvjat mladosti . Prolitje mladosti . Metaf.
Proljetno doba , illi proljetna .
Mladaz , tkomu ſmart nèmilla
<i>U proljetno bjesce doba</i>
Priko reda dni ſkratila . Palm. Chr. An-
ni della vecchiezza . <i>Anni ſeneſcentes</i> . Sta-
ra godifctaa . Anni della decrepitezza .
<i>Anni Acheruntici</i> . Viſcègrebna godifctaa .
Najpoſledgna godifctaa . Priſtara godifctaa .
<i>Akòsam priſtara , allisam priſtala</i> , Gun.
Dubr. Anni della puerizia . <i>Anni pueriles</i> .
Djetinska ljettaa . Djetigna godifctaa . Neſ-
rela doba .</p> | <p>disfeta . Vechje od godifceta . L' anno ap-
preſſo . <i>Post annum</i> . Drugo godifcte . Do-
hodnega godifceta . Do godifceta . Do ljet-
ta . L' anno paſſato . <i>Anno præterito</i> . La-
ni . Lansko godifcte . Proſcjaſto godifcte ,
alli ljetto . L' anno venturo . <i>Anno futu-
ro</i> . Dohodno godifcte . Due anni ſono .
<i>Duobus ab hinc annis</i> . Onomlani . Jeſu
dvje godine danaa . Jeſu dva ljetta . Na
ſad dva godifctaa . Anno temperato . <i>An-
nus temperatus</i> . Meko godifcte . Anno in-
ſalubre . <i>Annus inſalubris</i> . Neſdravo go-
difcte . Anni della diſcrezione , ſono quel-
li , ne' quali uno diſcerne il bene dal ma-
le . <i>Sapientia anni</i> . Raſloſcna godifctaa .
Spametna godifctaa . Raſborna godifctaa .
Avere gl' anni della diſcrezione , ovvero
l' uſo della ragione . <i>Sapio , is , pui , itum</i> .
Imati raſborna godifctaa . Mochi raſabrati
ſlo od dobra . Biti ſtupio na ljetaa od
raſbora . Anni giovanili , o della gioventù
.
<i>Anni crescentes</i> . Mlada godifctaa .
Zvjat mladosti . Prolitje mladosti . Metaf.
Proljetno doba , illi proljetna .
Mladaz , tkomu ſmart nèmilla
<i>U proljetno bjesce doba</i>
Priko reda dni ſkratila . Palm. Chr. An-
ni della vecchiezza . <i>Anni ſeneſcentes</i> . Sta-
ra godifctaa . Anni della decrepitezza .
<i>Anni Acheruntici</i> . Viſcègrebna godifctaa .
Najpoſledgna godifctaa . Priſtara godifctaa .
<i>Akòsam priſtara , allisam priſtala</i> , Gun.
Dubr. Anni della puerizia . <i>Anni pueriles</i> .
Djetinska ljettaa . Djetigna godifctaa . Neſ-
rela doba .</p> |
| <p>Anfiare , ed anfare , reſpirare con difficoltà .
<i>Anhelo , as ſapjehatiſe , pjehàmſe , hòsàm-
ſe</i> . Trudnòdihati , diham , hòſam . Pjeha-
ti , ham , hòſam . Stegnati , ſtegnem , ſte-
gnaòſam .</p> | <p>Anfiare , ed anfare , reſpirare con difficoltà .
<i>Anhelo , as ſapjehatiſe , pjehàmſe , hòsàm-
ſe</i> . Trudnòdihati , diham , hòſam . Pjeha-
ti , ham , hòſam . Stegnati , ſtegnem , ſte-
gnaòſam .</p> |
| <p>Aratro di più paja di buoi . <i>Aratrum mul-
tiplici jugo</i> . Plugh , ga . m.
<i>Er ſcto ufore ſa dan plugom</i> ,
<i>S' opchinòmùje hrane doſta</i> . Oſm. Cavar
qualche coſa coll' aratro . <i>Arando effodere</i> .
Iſorati , rajem , raòſam . Iſkopati , alli
iſnjeti ralom ſctògodi . Cavato coll' aratro .
<i>Aratro effoſus</i> . Iſoran , ana , ano . Il ferro
dell' aratro , cioè il vomero . <i>Vomer , ris</i> .
Lemiſc , ſcja . m. Raonik , ika . m. Orno
gvoſdje . Il dentale dell' aratro . <i>Plaſ , ſa</i> .
m. Manico dell' aratro . <i>Aratri manubrium</i> .
Raliza , ze . f. Darſcjak od rala .</p> | <p>Aratro di più paja di buoi . <i>Aratrum mul-
tiplici jugo</i> . Plugh , ga . m.
<i>Er ſcto ufore ſa dan plugom</i> ,
<i>S' opchinòmùje hrane doſta</i> . Oſm. Cavar
qualche coſa coll' aratro . <i>Arando effodere</i> .
Iſorati , rajem , raòſam . Iſkopati , alli
iſnjeti ralom ſctògodi . Cavato coll' aratro .
<i>Aratro effoſus</i> . Iſoran , ana , ano . Il ferro
dell' aratro , cioè il vomero . <i>Vomer , ris</i> .
Lemiſc , ſcja . m. Raonik , ika . m. Orno
gvoſdje . Il dentale dell' aratro . <i>Plaſ , ſa</i> .
m. Manico dell' aratro . <i>Aratri manubrium</i> .
Raliza , ze . f. Darſcjak od rala .</p> |
| <p>Arbitrato , add. coſa giudicata dall' Arbitro .
<i>Arbitratus , ta , tum</i> . Sudjen , na , no od
dobrjeh gljudji .</p> | <p>Arbitrato , add. coſa giudicata dall' Arbitro .
<i>Arbitratus , ta , tum</i> . Sudjen , na , no od
dobrjeh gljudji .</p> |
| <p>Arcajo , ed arcaro , facitor d' arche . <i>Arca-
rius , rii</i> . Skrignar , ara . m.</p> | <p>Arcajo , ed arcaro , facitor d' arche . <i>Arca-
rius , rii</i> . Skrignar , ara . m.</p> |

| | |
|--|--|
| Arrollare , scrivere a rotlo . <i>Conſcribo</i> , <i>bis</i> .
ſapisati , ſcjujem , ſaòſam . v. g. Vojnike
&c. ſbirati vojnike . Skupiti vojnike . Is-
pifati viteſove . | Arrollare , ſcrivere a rotlo . <i>Conſcribo</i> , <i>bis</i> .
ſapisati , ſcjujem , ſaòſam . v. g. Vojnike
&c. ſbirati vojnike . Skupiti vojnike . Is-
pifati viteſove . |
| <i>K' jagnicjarim' ter ſaviſce</i>
<i>Vele vechi broj iſ nova</i>
<i>Po iſboru joſc iſpiſce</i>
<i>Od Iſtoenjuh Viteſovaa . Oſm. Iſbirati</i>
<i>vojnike .</i>
<i>A i cette Spahogljanaa</i>
<i>Dvâkrat vechje s' tiem ucinj</i>
<i>S' nova ſkupa iſabrana . Oſm. Arrollarſi .</i>
<i>Dare nomen militiæ ſapitatife ſa vojnika .</i>
<i>Upifatife u vojsku .</i> | <i>K' jagnicjarim' ter ſaviſce</i>
<i>Vele vechi broj iſ nova</i>
<i>Po iſboru joſc iſpiſce</i>
<i>Od Iſtoenjuh Viteſovaa . Oſm. Iſbirati</i>
<i>vojnike .</i>
<i>A i cette Spahogljanaa</i>
<i>Dvâkrat vechje s' tiem ucinj</i>
<i>S' nova ſkupa iſabrana . Oſm. Arrollarſi .</i>
<i>Dare nomen militiæ ſapitatife ſa vojnika .</i>
<i>Upifatife u vojsku .</i> |
| Aſſaffinato . <i>Aprædatoribus deſpoliatus</i> . Po-
robgljen , na , no . Udart od guſe . | Aſſaffinato . <i>A prædatoribus deſpoliatus</i> . Po-
robgljen , na , no . Odart od guſe . |
| Attivo , che fà , o può fare . <i>Actuoſus</i> , <i>ſa</i> ,
<i>ſum</i> . Poſpjecjan u djellu , na , no . | Attivo , che fà , o può fare . <i>Actuoſus</i> , <i>ſa</i> ,
<i>ſum</i> . Poſpjecjan u djellu , na , no . |
| <i>Mnoſe na ſvako ſlo poſpjesna</i> . Palm. Chr.
Obartan , tna , tno . Djellòtvorni , na , no .
Otrefen , ena , eno . Hitar . Lak . Barſ u
poſlieh , u djellieh &c. | <i>Mnoſe na ſvako ſlo poſpjesna</i> . Palm. Chr.
Obartan , tna , tno . Djellòtvorni , na , no .
Otrefen , ena , eno . Hitar . Lak . Barſ u
poſlieh , u djellieh &c. |
| Aumentare , aumentarſi . V. Accreſcere , ac-
creſcerſi . | Aumentare , aumentarſi . V. Accreſcere , ac-
creſcerſi . |
| Avoltòio , uccello di rapina . <i>Vultur</i> , <i>riſ</i> . n.
Sep , pa . m. Koſtòſcer , era . m. Koſtòſc-
der , ra . m. Dobròhran , na . m.
<i>ſluſci ſa dom , i obranu</i>
<i>ſederglivomu Dobròhranu</i> . Gjorg. Pf. Di
avoltoio . <i>Vulturinus</i> , <i>a</i> , <i>um</i> . Koſtòſcer-
ni , na , no . Koſtòſcderni , na , no . | Avoltòio , uccello di rapina . <i>Vultur</i> , <i>riſ</i> . n.
Sep , pa . m. Koſtòſcer , era . m. Koſtòſc-
der , ra . m. Dobròhran , na . m.
<i>ſluſci ſa dom , i obranu</i>
<i>ſederglivomu Dobròhranu</i> . Gjorg. Pf. Di
avoltoio . <i>Vulturinus</i> , <i>a</i> , <i>um</i> . Koſtòſcer-
ni , na , no . Koſtòſcderni , na , no . |
| Avvenire , avv. per l' avvenire . <i>Poſthac</i> .
Unaprieda . Od ſada . | Avvenire , avv. per l' avvenire . <i>Poſthac</i> .
Unaprieda . Od ſada . |
| Bàmbagino , tela di bambagio . <i>Textile Goffi-
pium</i> . Tkani bumbak . Pamukni poſtav .
Di bambagio , <i>Goffipinus</i> , <i>a</i> , <i>um</i> . Bum-
bacni , na , no . Illir. Barb. Pamukni , na ,
no . uſato . | Bàmbagino , tela di bambagio . <i>Textile Goffi-
pium</i> . Tkani bumbak . Pamukni poſtav .
Di bambagio , <i>Goffipinus</i> , <i>a</i> , <i>um</i> . Bum-
bacni , na , no . Illir. Barb. Pamukni , na ,
no . uſato . |
| Banco , ſtomento di legno da federci più
perſone inſieme . <i>Subſellium</i> , <i>lii</i> . n. Klup ,
upa , m. Sjedaliſcte , ſcta . n. Sjedifcte ,
ſcta . n. Banco di galera , o altro naviglio .
<i>Tranſtrum</i> , <i>ſtri</i> . n. Plavski klup . Klup
od voſazaa . Illir. barb. klup od ghjemie .
Banco da due . <i>Biſellium</i> , <i>lii</i> . n. Klup ſa
dvoizu . Stoo ſa dvoize . Banco da Macel-
lajo . <i>Menſa lanionia</i> . Trup , pa . m. Tru-
pina , ne . f. | Banco , ſtomento di legno da federci più
perſone inſieme . <i>Subſellium</i> , <i>lii</i> . n. Klup ,
upa , m. Sjedaliſcte , ſcta . n. Sjedifcte ,
ſcta . n. Banco di galera , o altro naviglio .
<i>Tranſtrum</i> , <i>ſtri</i> . n. Plavski klup . Klup
od voſazaa . Illir. barb. klup od ghjemie .
Banco da due . <i>Biſellium</i> , <i>lii</i> . n. Klup ſa
dvoizu . Stoo ſa dvoize . Banco da Macel-
lajo . <i>Menſa lanionia</i> . Trup , pa . m. Tru-
pina , ne . f. |

Banda , cioè parte , o lato . *Pars , tis . f.*
 Strana , ne . f.
Alli Dervise desnu stranu
Hitro usmace Ofm. Banda
 dinnanzi . *Pars antica* . Spredgna strana .
 Sprjed , da . n. da ufarsi raramente . v. g.
 Moje sprjed ; in neutro . Od ovoga sprje-
 da . Dalla banda dinnanzi . *A fronte* . Sprje-
 da . S' sprjedgne strane . Banda di dietro .
Pars postica . fadgna strana . fad , od fada .
 n. si ufa però di rado . v. g. Moje fad .
 Dalla banda di dietro . *A tergo* . fada , s'
 fadgne strane . A banda sinistra . *Sinistror-*
sum . Na ljevu stranu . K' ljevoj strani . Na
 ljevu . A banda destra . *Dextrorsum* . Na

Banda , cioè parte , o lato . *Pars , tis . f.*
 Strana , ne . f.
Alli Dervise desnu stranu
Hitro usmace Ofm. Banda
 dinnanzi . *Pars antica* . Spredgna strana .
 Sprjed , da . n. da ufarsi raramente . v. g.
 Moje sprjed ; in neutro . Od ovoga sprje-
 da . Dalla banda dinnanzi . *A fronte* . Sprje-
 da . S' sprjedgne strane . Banda di dietro .
Pars postica . fadgna strana . fad , od fada .
 n. si ufa però di rado . v. g. Moje fad .
 Dalla banda di dietro . *A tergo* . fada , s'
 fadgne strane . A banda sinistra . *Sinistror-*
sum . Na ljevu stranu . K' ljevoj strani . Na
 ljevu . A banda destra . *Dextrorsum* . Na

desnu stranu . K' desnoj strani . Na desnu .
 Banda di dentro . *Pars interior* . Unutar-
 gna strana . Unutra . v. g. Moje unutra ;
 in neutro . Banda di fuori . *Pars exterior* .
 Isvagna strana . Dvorgna strana . fdivora .
 v. g. Moje fdivora . Banda di sotto . *Pars*
inferior . Dogna strana . Strana odijdo ,
 alliti o/dala . Banda di sopra . *Pars supe-*
rior . Gorgna strana . Strana od fgori , al-
 liti odigara . Da qualche banda . *Alicun-*
de . Od kègod strane . Od kudàgodi . Od
 klègodi . Da ogni banda . *Undique* . Od
 fvae strane . Od fvjeh stranaa . Sa fvjeh
 stranaa . Odàfvuda . Od fvdjere .
A to , da tuj ne pafechi
Sipat pjenef od svudjere . Ofm. Da ban-
 da a banda , come passàr da banda a ban-
 ba . *Transverbero , as , avi , atum* . Probiti
 s' strane na stranu , biam , biòfam .
Pribjenèsu cjavlìm k' darvu
S' ove strane na onu stranu . Mand. Bun.
 Prodjefti tkoga kopjem , macem &c. dje-
 vam , diòfam .
Sprjèdmu ifide scip prof plechja . Ofm.
 Da banda , da parte . *Seorsim* . Pò nase .
 S' strane . Na stranu . Na pònase . Na pò-
 se . Pòse . Banda , traversa di drap-
 po , che si porta al colo . *Fascia collaris* .
 Nadrameniza , ze . f. Prikorameniza , ze .
 f. Banda , compagnia de soldati . *Cohors* ,
tis . f. Cetta , te . f. Banda di cavalli , o
 di soldati a cavallo . *Turma equitum* . Cet-
 ta kognikaa , illiti kognicka cetta . Da
 ogni banda , da ogni lato , da ogni parte .
Quoquoversus . Od fvjeh krajaa . Od fvjeh
 stranaa . Oda fvuda .

desnu stranu . K' desnoj strani . Na desnu .
 Banda di dentro . *Pars interior* . Unutar-
 gna strana . Unutra . v. g. Moje unutra ;
 in neutro . Banda di fuori . *Pars exterior* .
 Isvagna strana . Dvorgna strana . fdivora .
 v. g. Moje fdivora . Banda di sotto . *Pars*
inferior . Dogna strana . Strana odijdo ,
 alliti o/dala . Banda di sopra . *Pars supe-*
rior . Gorgna strana . Strana od fgori , al-
 liti odigara . Da qualche banda . *Alicun-*
de . Od kègod strane . Od kudàgodi . Od
 klègodi . Da ogni banda . *Undique* . Od
 fvae strane . Od fvjeh stranaa . Sa fvjeh
 stranaa . Odàfvuda . Od fvdjere .
A to , da tuj ne pafechi
Sipat pjenef od svudjere . Ofm. Da ban-
 da a banda , come passàr da banda a ban-
 ba . *Transverbero , as , avi , atum* . Probiti
 s' strane na stranu , biam , biòfam .
Pribjenèsu cjavlìm k' darvu
S' ove strane na onu stranu . Mand. Bun.
 Prodjefti tkoga kopjem , macem &c. dje-
 vam , diòfam .
Sprjèdmu ifide scip prof plechja . Ofm.
 Da banda , da parte . *Seorsim* . Pò nase .
 S' strane . Na stranu . Na pònase . Na pò-
 se . Pòse . Banda , traversa di drap-
 po , che si porta al colo . *Fascia collaris* .
 Nadrameniza , ze . f. Prikorameniza , ze .
 f. Banda , compagnia de soldati . *Cohors* ,
tis . f. Cetta , te . f. Banda di cavalli , o
 di soldati a cavallo . *Turma equitum* . Cet-
 ta kognikaa , illiti kognicka cetta . Da
 ogni banda , da ogni lato , da ogni parte .
Quoquoversus . Od fvjeh krajaa . Od fvjeh
 stranaa . Oda fvuda .

| | |
|---|---|
| Beneficio , grazia , piacere . <i>Beneficium</i> , <i>cii</i> . n. Gljubav , vi . f. Millos , ſti . f. Har , ri . f. <i>Er ſzjenecz</i> , <i>da ſa har imatiche ſcto cinih</i> . Scifc. Dobrodarje , rja . n. Dobročinstvo , ſtva , n. Render beneficio . <i>Beneficium red-dere</i> . Vratiti gljubav . Vratiti har , alliti millos . Beneficio , entrata , o rendita cle-ricale . <i>Vectigal Eccleſiaſticum</i> . Zarkovni do-hodak . Nadarje Zarkovno . Djacija Zarkov-na . Beneficio ſemplice . <i>Sacerdotium curio-natu vacuum</i> . T. Eccl. Zarkovni dohodak beſ pomgne od duſcjaa . Slobodna djacija . Beneficio curato , che ha cura d' Anime . <i>Sacerdotium Animarum cura affectum</i> . T. Eccl. Zarkovno nadarje s' pomgnom od duſcjaa . Djacija s' vladagnem od Duſcjaa . djacija duſcjòvladna , illi Vladòduſcja . Be-neficio di molta rendita . <i>Sacerdotium pin-gue</i> . T. Eccl. Nadarje Zarkovnu bogato . Djacija bogata . Beneficio piccolo , di poca rendita . <i>Sacerdotium tenue</i> . T. Eccl. Nada-rje Zarkovno mallòkorifno . Mallòkorifna Djacija . Uboga Djacija . | Beneficio , grazia , piacere . <i>Beneficium</i> , <i>cii</i> . n. Gljubav , vi . f. Millos , ſti . f. Har , ri . f. <i>Er ſzjenecz</i> , <i>da ſa har imatiche ſcto cinih</i> . Scifc. Dobrodarje , rja . n. Dobročinstvo , ſtva , n. Render beneficio . <i>Beneficium red-dere</i> . Vratiti gljubav . Vratiti har , alliti millos . Beneficio , entrata , o rendita cle-ricale . <i>Vectigal Eccleſiaſticum</i> . Zarkovni do-hodak . Nadarje Zarkovno . Djacija Zarkov-na . Beneficio ſemplice . <i>Sacerdotium curio-natu vacuum</i> . T. Eccl. Zarkovni dohodak beſ pomgne od duſcjaa . Slobodna djacija . Beneficio curato , che ha cura d' Anime . <i>Sacerdotium Animarum cura affectum</i> . T. Eccl. Zarkovno nadarje s' pomgnom od duſcjaa . Djacija s' vladagnem od Duſcjaa . djacija duſcjòvladna , illi Vladòduſcja . Be-neficio di molta rendita . <i>Sacerdotium pin-gue</i> . T. Eccl. Nadarje Zarkovnu bogato . Djacija bogata . Beneficio piccolo , di poca rendita . <i>Sacerdotium tenue</i> . T. Eccl. Nada-rje Zarkovno mallòkorifno . Mallòkorifna Djacija . Uboga Djacija . |
| Bigamia . <i>Bigamia</i> , <i>mie</i> . Dvòſcegnegne , gna . n. | Bigamia . <i>Bigamia</i> , <i>mie</i> . Dvòſcegnegne , gna . n. |
| Borraccia , fiaſco di cuojo da tener vino , o altro liquore . <i>Scortea</i> , <i>tea</i> . f. Buklia , e . f. Matara , re . f. | Borraccia , fiaſco di cuojo da tener vino , o altro liquore . <i>Scortea</i> , <i>tea</i> . f. Buklia , e . f. Matara , re . f. |
| Bottigliere , colui , che tiene cura del vino , che ſi dà a tavola . <i>Vini epularis curator</i> . Nadpivnik , ka . m. Kgljucjar od vinize . Kuchni karcmar . <i>Kgljucemi oncjaa da od karcme Bolkovich</i> : <i>Kruh</i> , <i>larad</i> , <i>vladam svim</i> , <i>a kupgljuse u vinu</i> . Darſc. Raſl. | Bottigliere , colui , che tiene cura del vino , che ſi dà a tavola . <i>Vini epularis curator</i> . Nadpivnik , ka . m. Kgljucjar od vinize . Kuchni karcmar . <i>Kgljucemi oncjaa da od karcme Bolkovich</i> : <i>Kruh</i> , <i>larad</i> , <i>vladam svim</i> , <i>a kupgljuse u vinu</i> . Darſc. Raſl. |
| Broccoli , cime di cavoli più tenere , e deli-cate . <i>Brassicarum cymæ</i> . f. pl . Prokoliza , ze . f. Klize kupuſne . f. pl. Prokuliza . | Broccoli , cime di cavoli più tenere , e deli-cate . <i>Brassicarum cymæ</i> . f. pl . Prokoliza , ze . f. Klize kupuſne . f. pl. Prokuliza . |
| Bufalo . <i>Bubalus</i> , <i>li</i> . m. Bivoo , la . m. <i>Bivoli od ſprjed</i> , <i>kji do paſa Cjovjecjanske ſlike noſe</i> . Oſm. Carne di bufalo . <i>Caro bubalina</i> . Bivoluie meſo . Bi-volsko meſo . | Bufalo . <i>Bubalus</i> , <i>li</i> . m. Bivoo , la . m. <i>Bivoli od ſprjed</i> , <i>kji do paſa Cjovjecjanske ſlike noſe</i> . Oſm. Carne di bufalo . <i>Caro bubalina</i> . Bivoluie meſo . Bi-volsko meſo . |
| Bugiare , dir bugia , mentire . <i>Mentior</i> , <i>iris</i> . Lagati , laſcem , lagaòſam . Slagati , laſcem , slagaòſam . Svoditi laſc , ſvodim , ſvodio-ſam laſc . <i>Njednùt' laſc ne ſvodim &c.</i> Elekt. | Bugiare , dir bugia , mentire . <i>Mentior</i> , <i>iris</i> . Lagati , laſcem , lagaòſam . Slagati , laſcem , slagaòſam . Svoditi laſc , ſvodim , ſvodio-ſam laſc . <i>Njednùt' laſc ne ſvodim &c.</i> Elekt. |
| Capria , capivola , capra ſalvatica . <i>Caprea</i> , æ . f. Sarna , ne . f. | Capria , capivola , capra ſalvatica . <i>Caprea</i> , æ . f. Sarna , ne . f. |
| Ceſto . V. Canettro . | Ceſto . V. Canettro . |

| | |
|---|---|
| Chiarire , render chiaro cofa , o negozio . <i>Dilucidare explicare</i> . Rafabrati , birem , brad-
fam . Rafabitrati , trivam , triòfam . Occito
prika/fati , &c. Chiarire uno , diceli di chi
fa qualche burla ad un altro . <i>Imponere a-
licui</i> . Skroiti tkomu , skroim , skroiòfam .
Metaf. Chiarirfi di qualche cofa . <i>Certum
aliquid fcire</i> . Dofvjettitife , tujèmfe , tiò-
sàmfe . Dochi u po/agne . V. Avvederfi . | Chiarire , render chiaro cofa , o negozio .
<i>Dilucidare explicare</i> . Rafabrati , birem , brad-
fam . Rafabitrati , trivam , triòfam . Occito
prika/fati , &c. Chiarire uno , diceli di chi
fa qualche burla ad un altro . <i>Imponere a-
licui</i> . Skroiti tkomu , skroim , skroiòfam .
Metaf. Chiarirfi di qualche cofa . <i>Certum
aliquid fcire</i> . Dofvjettitife , tujèmfe , tiò-
sàmfe . Dochi u po/agne . V. Avvederfi . |
| Chirógrafo . <i>Chirographus</i> . Rukòpifagne , gna .
n. Rukòpis , fa . m. | Chirógrafo . <i>Chirographus</i> . Rukòpifagne , gna .
n. Rukòpis , fa . m. |
| Chiunque . V. Qualunque . | Chiunque . V. Qualunque . |
| Coagulare . V. Quagliare . | Coagulare . V. Quagliare . |
| Coeguale , add. infieme eguale . <i>Coaqualis</i> ,
& hoc le . n. Jednòvarftan , na , no . | Coeguale , add. infieme eguale . <i>Coaqualis</i> ,
& hoc le . n. Jednòvarftan , na , no . |
| Coetanea . <i>Coæva</i> , æ . Varfniza , ze . f. | Coetanea . <i>Coæva</i> , æ . Varfniza , ze . f. |
| Cóltrice , feconda breve . V. Matarazzo . | Cóltrice , feconda breve . V. Matarazzo . |
| Commettere , dar commiffione . <i>Demando</i> , as .
Narediti , djujem , diòfam . Odrediti , dju-
jem , diòfam . Priporuciti , cjujem , ciòfam .
Nasloniti tkomu , naslagnam , niòfam . fa-
povidjeti , dam , diòfam . Freq. Narediva-
ti . Priporucivati . Naslonivati tkomu pofo .
Commettere errore . V. Errare . Commit-
tere , fare . V. Fare . Commettere , inca-
ftrare . V. Incaftrare . | Commettere , dar commiffione . <i>Demando</i> , as .
Narediti , djujem , diòfam . Odrediti , dju-
jem , diòfam . Priporuciti , cjujem , ciòfam .
Nasloniti tkomu , naslagnam , niòfam . fa-
povidjeti , dam , diòfam . Freq. Narediva-
ti . Priporucivati . Naslonivati tkomu pofo .
Commettere errore . V. Errare . Commit-
tere , fare . V. Fare . Commettere , inca-
ftrare . V. Incaftrare . |
| Cómputo . V. Abbaco . | Cómputo . V. Abbaco . |
| Comunque , in qualunque modo . <i>Utrumque</i> .
Kakòmùdrago . Na koimùdrago nacin . Co-
munque fía . <i>Ut ut eft</i> . Kakòmùdrago . | Comunque , in qualunque modo . <i>Utrumque</i> .
Kakòmùdrago . Na koimùdrago nacin . Co-
munque fía . <i>Ut ut eft</i> . Kakòmùdrago . |
| Cóncia , luogo , dove fi conciano le pelli .
<i>Coriarii officina</i> . Tabakaria , rie . f. | Cóncia , luogo , dove fi conciano le pelli .
<i>Coriarii officina</i> . Tabakaria , rie . f. |
| Conclufióne , termine del difcorfo . <i>Conclu-
fio</i> , nis . f. faglava , ve , f. | Conclufióne , termine del difcorfo . <i>Conclu-
fio</i> , nis . f. faglava , ve , f. |
| Na faglavu tjeħ besjedaa &c. Palm. Chr.
faglavgljegne , gna . n. Svarha , he . f. Do-
varfcegne , gna . n. favarfcja , e . f. Con-
clufione , diceli ad ogni propofizione di
quelle , che fi pubblicano per difputare .
<i>Positio</i> , nis . f. faglavak , vka . m. Soften-
tar , o difender la conclufione . <i>Thefer propu-
gnare</i> . Braniti faglavke . Softentor , o di-
fenfor delle conclufioni . <i>Propugnator the-
fium</i> . Branitegl faglavakaa . | Na faglavu tjeħ besjedaa &c. Palm. Chr.
Jaglavgljegne , gna . n. Svarha , he . f. Do-
varfcegne , gna . n. favarfcja , e . f. Con-
clufione , diceli ad ogni propofizione di
quelle , che fi pubblicano per difputare .
<i>Positio</i> , nis . f. faglavak , vka . m. Soften-
tar , o difender la conclufione . <i>Thefes propu-
gnare</i> . Braniti faglavke . Softentor , o di-
fenfor delle conclufioni . <i>Propugnator the-
fium</i> . Eranitegl faglavakaa . |
| Conquaffatrice . <i>Conquaffatrix</i> , cis . f. Rafu-
fcitegliza , ze . f. Satritegliza , ze . f. Sa-
tarifaliza , ze . f. | Conquaffatrice . <i>Conquaffatrix</i> , cis . f. Rafu-
fcitegliza , ze . f. Satritegliza , ze . f. Sa-
tarifaliza , ze . f. |

| | |
|---|---|
| Contrário , fuft. <i>Vir adversus</i> . Nepriategl , glja . m. Protivnik , ka . m. Suprotivnik , ka . m. Per lo contrario , avv. ovvero all' incontro . <i>Contra</i> . S' drughe frane . | Contrário , fuft. <i>Vir adversus</i> . Nepriategl , glja . m. Protivnik , ka . m. Suprotivnik , ka . m. Per lo contrario , avv. ovvero all' incontro . <i>Contra</i> . S' drughe frane . |
| Convertito . <i>Conversus</i> , a , um . Obrachjen , na , no . Priveden na dobar put . Obarnut , ta , to . Svarnut , ta , to . | Convertito . <i>Conversus</i> , a , um . Obrachjen , na , no . Priveden na dobar put . Obarnut , ta , to . Svarnut , ta , to . |
| Corallo , pianta , che nafce nel fondo del mare . <i>Corallium</i> , lii . n. Koragl , glja . m. Od koraglja usti objavj , A od lira parsi svoje , Rafm. Di corallo . <i>Corallinus</i> , a , um . Koraglni , na , no . Koragline dvje usti &c. Gljub. Prof koraglna neka vrata . Na dvor fejarko Sunze ifide . Ofm. | Corallo , pianta , che nafce nel fondo del mare . <i>Corallium</i> , lii . n. Koragl , glja . m. Od koraglja usti objavj , A od lira parsi svoje , Rafm. Di corallo . <i>Corallinus</i> , a , um . Koraglni , na , no . Koragline dvje usti &c. Gljub. Prof koraglna neka vrata . Na dvor fejarko Sunze ifide . Ofm. |
| Corbacchiotto , corbicino , il figlio del corvo . <i>Corvi pullus</i> . Gravranich , chja . m. Vranich , chja . m. | Corbacchiotto , corbicino , il figlio del corvo . <i>Corvi pullus</i> . Gravranich , chja . m. Vranich , chja . m. |
| Cordiale , cioè intimo . <i>Intimus</i> , a , um . Sarcjani , na , no . | Cordiale , cioè intimo . <i>Intimus</i> , a , um . Sarcjani , na , no . |
| Corrente , fuft. acqua , che corre . V. Corfia . Corrente , furiofo . <i>Torrentis</i> , tis . m. Poplaviza , ze . f. Tako kàdse nif brjegh starmi Poplaviza plaha obarne . Palm. Chr. Con la corrente , fecondo la corrente . <i>Secundo flumine</i> . S' vodom . Nif vodu . Contro la corrente , contro l' acqua . <i>Amne contrario</i> . Suproch vodi . Uf vodu . | Corrente , fuft. acqua , che corre . V. Corfia . Corrente , furiofo . <i>Torrentis</i> , tis . m. Poplaviza , ze . f. Tako kàdse nif brjegh starmi Poplaviza plaha obarne . Palm. Chr. Con la corrente , fecondo la corrente . <i>Secundo flumine</i> . S' vodom . Nif vodu . Contro la corrente , contro l' acqua . <i>Amne contrario</i> . Suproch vodi . Uf vodu . |
| Corvo . V. Corbo . | Corvo , V. Corbo . |
| Cofcienza . V. Confcienza . | Cofcienza . V. Confcienza . |
| Coftipáto nel detto fenfo . <i>Alvus conftipatus</i> . Tjelo fatvoreno . | Coftipáto nel detto fenfo . <i>Alvus conftipatus</i> . Tjelo fatvoreno . |
| Cótica , pelle dura dell' animale . V. Cuojo . | Cótica , pelle dura dell' animale . V. Cuojo . |
| Cozzóne , colui che doma i cavalli . <i>Domitor eorum</i> . Kognicjar , ara . m. | Cozzóne , colui che doma i cavalli . <i>Domitor eorum</i> . Kognicjar , ara . m. |
| Credo , il fimbo'lo degli Apostoli , o della Fede . <i>Apostolorum symbolum</i> . T. Eccl. Vje-rovagne , gna . n. | Credo , il fimbo'lo degli Apostoli , o della Fede . <i>Apostolorum symbolum</i> . T. Eccl. Vje-rovagne , gna . n. |

| | |
|---|---|
| <p>Cristo , Dio , ed uomo insieme , e Salvator del mondo . <i>Christus , sti</i> . Ifukars , Ifukarsta . m. Cristo crocifisso . <i>Christus cruce pendens</i> . Ifukars propeti . Defcr .</p> <p>Bjesce Spasitegl od svjeta raspct , i rasteghnut na kriscju , goo , nagh , isderan , israghen . Tezjasce karv Boscja s' glave po kosah . po bradi , po svemu obrafu . Tochjahuse is nogaa , is rukaa potozi privrjedne karvi , i nèimasce ghdi nasloniti plemenitu glavu dracjami okrugnenu . Kafs . If . Cristo risuscitato . <i>Christus redivivus</i> . Ifukars uskarfnut .</p> | <p>Cristo , Dio , ed uomo insieme , e Salvator del mondo . <i>Christus , sti</i> . Ifukars , Ifukarsta . m. Cristo crocifisso . <i>Christus cruce pendens</i> . Ifukars propeti . Defcr .</p> <p>Bjesce Spasitegl od svjeta raspct , i rasteghnut na kriscju , goo , nagh , isderan , israghen . Tezjasce karv Boscja s' glave po kosah . po bradi , po svemu obrafu . Tochjahuse is nogaa , is rukaa potozi privrjedne karvi , i nèimasce ghdi nasloniti plemenitu glavu dracjami okrugnenu . Kafs . If . Cristo risuscitato . <i>Christus redivivus</i> . Ifukars uskarfnut .</p> |
| <p>Crocetta , croce piccola . <i>Crux parva</i> . Kri-fcjak , fcka . m. Kri/cich , fcichja . m.</p> | <p>Crocetta , croce piccola . <i>Crux parva</i> . Kri-fcjak , fcka . m. Kri/cich , fcichja . m.</p> |
| <p>Croftàta , forte di torta , che si fa di frutti cotti con zucchero tra sottilissime croste di pane . <i>Placenta crustata</i> . Slatka kruftaviza . Prifnaz , od prifnaza . m. propriamente torta col formaggio .</p> | <p>Koriza , ze . f .</p> <p>Croftàta , forte di torta , che si fa di frutti cotti con zucchero tra sottilissime croste di pane . <i>Placenta crustata</i> . Slatka kruftaviza . Prifnaz , od prifnaza . m. propriamente torta col formaggio .</p> |
| <p>Cubo , corpo quadrato di fci lati eguali . <i>Cubus , bi</i> . m. Scefterònuhla , ghle . f.</p> | <p>Cubo , corpo quadrato di fci lati eguali . <i>Cubus , bi</i> . m. Scefterònuhla , ghle . f.</p> |
| <p>Curiófo , add. che ha curiosità . <i>Curiosus , a , um</i> . fcegljan , eglua , eglno . Naghli , gla , glo .</p> <p>Cim lipote draghe toli</p> <p>Naghlím okkom svaki kupi . Ofm . Naglò-fcegljan , glua , glno . Curiofo , chi tien cura , o si prende pensiero di qualche cosa . <i>Curiosus , a , um</i> . Pomnechi , chja , chje . Pomnik , ka . m. Poman , mna , mno . Curiofo di sapere . <i>Sciendi cupidus</i> . fceglno/nanaz , nza . m. fcegljan fnati .</p> | <p>Curiófo , add. che ha curiosità . <i>Curiosus , a , um</i> . fcegljan , eglua , eglno . Naghli , gla , glo .</p> <p>Cim lipote draghe toli</p> <p>Naghlím okkom svaki kupi . Ofm . Naglò-fcegljan , glua , glno . Curiofo , chi tien cura , o si prende pensiero di qualche cosa . <i>Curiosus , a , um</i> . Pomnechi , chja , chje . Pomnik , ka . m. Poman , mna , mno . Curiofo di sapere . <i>Sciendi cupidus</i> . fceglno/nanaz , nza . m. fcegljan fnati .</p> |
| <p>Danneggiatrice . <i>Damnifica</i> . Ofctetitegliza , ze . f.</p> | <p>Danneggiatrice . <i>Damnifica</i> . Ofctetitegliza , ze . f.</p> |
| <p>Decina , quantità , o numero di dieci . <i>Decas , dis</i> . f. Defetero . Defetak . Defetina , ne . f. Mezza decina . <i>Pentas , dis</i> . f. Petero . Pet . Petina , ne . f.</p> | <p>Decina , quantità , o numero di dieci . <i>Decas , dis</i> . f. Defetero . Defetak . Defetina , ne . f. Mezza decina . <i>Pentas , dis</i> . f. Petero . Pet . Petina , ne . f.</p> |
| <p>Degradàre , privare della dignità sacerdotale . <i>Sacerdotali gradu exuere</i> . Potifnuti s' Redovnistva . Obaliti s' visine doftojanstva Redovnickoga .</p> | <p>Degradàre , privare della dignità sacerdotale . <i>Sacerdotali gradu exuere</i> . Potifnuti s' Redovnistva . Obaliti s' visine doftojanstva Redovnickoga .</p> |
| <p>Diamantino , fatto di diamante . <i>Adamantinus , a , um</i> . Od diamanta .</p> | <p>Diamantino , fatto di diamante . <i>Adamantinus , a , um</i> . Od diamanta .</p> |
| <p>Dibattuto , disputato . <i>Discussus , a , um</i> . Pro-trefen , na , no . Ra/fcefcgljan , na , no .</p> | <p>Dibattuto , disputato . <i>Discussus , a , um</i> . Pro-trefen , na , no . Ra/fcefcgljan , na , no .</p> |
| <p>Diffidenza , disconfidenza . <i>Diffidentia , tia</i> . f. Nèu/dagne , gna . n. Neufagne , gna . n. Nèu/danstvo , stva . n. Nèpria/nos , sti . f.</p> | <p>Diffidenza , disconfidenza . <i>Diffidentia , tia</i> . f. Nèu/dagne , gna . n. Neufagne , gna . n. Nèu/danstvo , stva . n. Nèpria/nos , sti . f.</p> |

| | |
|--|---|
| Dignità , cioè grado d' ufficio , o preeminen-
za . V. Degnità . | Dignità , cioè grado d' ufficio , o preeminen-
za . V. Degnità . |
| Dimagràre , imagrire . V. Smagrarè . | Dimagràre , imagrire . V. Smagrarè . |
| <p>Dire , vale esprimere con parole . <i>Dico , cis .</i>
 Rechi , illi rjeti , govorim , rekaòfam . Go-
 voriti , rim , riòfam . Praviti , vim , viòfam .
 <i>On facjudjen pravi ; Tili</i>
 <i>Slufi tvomu peresc noghe ?</i> Palm. Chr.
 <i>I gnegov tac fcestok ogàgnse svjem pravi ,</i>
 Ragn. Rafl. Klikovati , kujem , kovaòfam .
 <i>Pak u sebi klikovati</i>
 <i>Pocce ; ovòsu svete noghe ,</i> Mand. Bun.
 Veliti , lim , liòfam . Ifultiti , tjujem , tiòfam .
 <i>Jesus , i ove rjeci ifustj .</i> Palm. Chr. Be-
 sjediti , besjedim , diòfam .
 <i>Ter pogledim' veselime</i>
 <i>Tufcnjem rjeci sej besjedi .</i> Palm. Chr.
 Ifrizati , zam besjede , ifrizòfam .
 <i>I fcejalosno ifrizati</i>
 <i>Milloj drufcbi sej besjede .</i> Palm. Chr.
 Pufititi besjede .
 <i>I medenjuh kada ustji ,</i>
 <i>If kjéh rajska rados grede ,</i>
 <i>Put Nebesa Jesus pustj</i>
 <i>Vjecnom' Otzu sej besjede ;</i> Palm. Chr.
 Oglafitife u rjeci &c.
 <i>U ovakèse tufcbe oglasj ,</i> Osm. Dire , rac-
 contare . V. Raccontare . Altro dire , al-
 tro fentire , altrimenti penfare , altrimen-
 ti parlare . <i>Alia loqui , alia fentire .</i> Jedna
 govoriti , druga chjutiti . Nèodgovarati je-
 fik farzu . Dir tutto . <i>Totum dicere .</i> Ifgo-
 voriti , varam , riòfam . Ifrjeti , ifrecem ,
 ifrekòfam . Dire brevemente . <i>Strictum di-</i>
 <i>cere .</i> Rjeti u kratko . Rjeti u jednu rjec .
 U mallo rjecji . Dire a bocca . <i>Coram lo-</i>
 <i>qui .</i> Rechi na okkòrjec . Rechi na rjec .
 Dire a capriccio , parlare inconfideratamen-
 te . <i>Inconfiderate loqui .</i> Govoriti fcto do-
 ghje na ufta . Govoriti nèra/borito . Spar-
 dati , dam , spardaòfam . Cepgljuskati ,
 kam , kòfam . L' una , e l' altra . V. Bafs.
 V. Ciarlare . Dire a lettere di scattola ,
 cioè chiaramente . <i>Clare dicere .</i> Oblo go-
 voriti . Govoriti efckero . Occito rechi . Go-
 voriti bjelòdano . Dire alla sfuggita . <i>Cur-</i>
 <i>sim dicere .</i> Grede govoriti . Hode rechi
 fctògod . Ufgrede rjeti . Dire in secreto .</p> | <p>Dire , vale esprimere con parole . <i>Dico , cis .</i>
 Rechi , illi rjeti , govorim , rekaòfam . Go-
 voriti , rim , riòfam . Praviti , vim , viòfam .
 <i>On facjudjen pravi ; Tili</i>
 <i>Slufi tvomu peresc noghe ?</i> Palm. Chr.
 <i>I gnegov tac fcestok ogàgnse svjem pravi ,</i>
 Ragn. Rafl. Klikovati , kujem , kovaòfam .
 <i>Pak u sebi klikovati</i>
 <i>Pocce ; ovòsu svete noghe ,</i> Mand. Bun.
 Veliti , lim , liòfam . Ifultiti , tjujem , tiòfam .
 <i>Jesus , i ove rjeci ifustj .</i> Palm. Chr. Be-
 sjediti , besjedim , diòfam .
 <i>Ter pogledim' veselime</i>
 <i>Tufcnjem rjeci sej besjedi .</i> Palm. Chr.
 Ifrizati , zam besjede , ifrizòfam .
 <i>I fcejalosno ifrizati</i>
 <i>Milloj drufcbi sej besjede .</i> Palm. Chr.
 Pufititi besjede .
 <i>If medenjuh kada ustji ,</i>
 <i>If kjéh rajska rados grede ,</i>
 <i>Put Nebesa Jesus pustj</i>
 <i>Vjecnom' Otzu sej besjede ;</i> Palm. Chr.
 Oglafitife u rjeci &c.
 <i>U ovakèse tufcbe oglasj ,</i> Osm. Dire , rac-
 contare . V. Raccontare . Altro dire , al-
 tro fentire , altrimenti penfare , altrimen-
 ti parlare . <i>Alia loqui , alia fentire .</i> Jedna
 govoriti , druga chjutiti . Nèodgovarati je-
 fik farzu . Dir tutto . <i>Totum dicere .</i> Ifgo-
 voriti , varam , riòfam . Ifrjeti , ifrecem ,
 ifrekòfam . Dire brevemente . <i>Strictum di-</i>
 <i>cere .</i> Rjeti u kratko . Rjeti u jednu rjec .
 U mallo rjecji . Dire a bocca . <i>Coram lo-</i>
 <i>qui .</i> Rechi na okkòrjec . Rechi na rjec .
 Dire a capriccio , parlare inconfideratamen-
 te . <i>Inconfiderate loqui .</i> Govoriti fcto do-
 ghje na ufta . Govoriti nèra/borito . Spar-
 dati , dam , spardaòfam . Cepgljuskati ,
 kam , kòfam . L' una , e l' altra . V. Bafs.
 V. Ciarlare . Dire a lettere di scattola ,
 cioè chiaramente . <i>Clare dicere .</i> Oblo go-
 voriti . Govoriti efckero . Occito rechi . Go-
 voriti bjelòdano . Dire alla sfuggita . <i>Cur-</i>
 <i>sim dicere .</i> Grede govoriti . Hode rechi
 fctògod . Ufgrede rjeti . Dire in secreto .</p> |

| | |
|---|---|
| <p><i>Secreto dicere</i> . Dorizati potajno , alli u potaj .
 <i>Jedan drugom' vites s' parva</i> .
 <i>Ovi u potaj glas doriza</i> . Rechi potajno .
 Dir la bugia . Rjeti lafc . Lagati . Dir di si . Ajo , is . Veliti , lim , liòfam , alli rjeti takòje . Dir di no . Nego , as . Njekati , kam , kaòfam . fanjekati , kam , kòfam . Rjeti , nije tako &c. Dir francamente .
 <i>Præfenti animo dicere</i> . Slobodno govoriti .
 Dire all' improvviso . <i>Ex tempore dicere</i> .
 Iſnènade beſjediti . Nènadno govoriti . Rechi nèmiſliuſci . Dir male d' alcuno . <i>Male-dice de aliquo dicere</i> ſlo o tkomu govoriti . ſlòrechi od tkoga , recem , kaòfam . Koſiti tkoga jeſikom . V. Detrarre . Chi dice mal d' alcuno . <i>Obloquutor , ris . m.</i> ſlòrek , ka . m. ſlògovornik , ka . m. ſlògovoraz , za . m. Il femm. ſlògovorniza , ze .
 Dir all' orecchio . <i>In aure dicere</i> . Poſcjap-tati , ſcjapchjem , taòfam . Scjaptati , chjem , taòfam . Govoriti na uho . Freq. Scjapchi-vati , vam , vòfam . Poſcjapchivati , vam , vòfam . Dire a propoſito . <i>Apte dicere</i> .
 Sputno govoriti . Skladno , podobno , po putu govoriti . Dire a ſpropoſito . <i>Inepte dicere</i> . Nèſputno beſjediti . Nèskladno , nè-podobno govoriti . A dir il vero . <i>Ut ve-rum dicam</i> ſa rechi pravo . ſa rechi iſti-nu . Dico il vero . <i>Verum dico</i> . Pravvo dim . Govorim iſtinu . Dir di cuore . <i>Ex animo dicere</i> . Govoriti od farza . Govoriti iſ far-za . Dir da dovero . <i>Serio dicere</i> . Govoriti na sbigl . Dire a mente . <i>Dicere memoriter</i> . Rechi na pamet . Dire leggendo . <i>Dicere de ſcripto</i> . Rechi iſ kgnighe . Dir non vo-lendo , dir una parola per un' altra . <i>Labi lingua</i> . Utechi Jeſik . ſarechiſe , ricèmſe , ſarekaòſàmſe . Dir l' iſteſſo . <i>Cantilenam candem canere</i> . Pjevati iſtu pjeſan . Sve jedno govoriti . Dir pro , e contra a qual-cuno . <i>In utramque partem dicere</i> ſa tko-ga , i na tkoga rechi . Dir la coſa come ſta . <i>Rem ut eſt eloqui</i> . Rjeti onako , kakò-je . Dico a te . <i>Tecum loquor</i> . Tebi govo-rim . S' tobom govorim . Queſto ti dico . <i>Hoc ajo</i> . Ovòti velim . Ovo dim .
 <i>Dim Gascpara Milloſtičja</i> ; Ofm. Si dice . <i>Fama eſt</i> . Glas jeſt . Govoriſe . Per dir me-glio . <i>Ut verius dicam</i> ſa boglje rjeti . Non ſia detto per arroganza . <i>Abſit verbo inju-riæ</i> . Nèbud' receno ſa okolaſt .</p> | <p><i>Secreto dicere</i> . Dorizati potajno , alli u potaj .
 <i>Jedan drugom' vites s' parva</i> .
 <i>Ovi u potaj glas doriza</i> . Rechi potajno .
 Dir la bugia . Rjeti lafc . Lagati . Dir di si . Ajo , is . Veliti , lim , liòfam , alli rjeti takòje . Dir di no . Nego , as . Njekati , kam , kaòfam . fanjekati , kam , kòfam . Rjeti , nije tako &c. Dir francamente .
 <i>Præfenti animo dicere</i> . Slobodno govoriti .
 Dire all' improvviso . <i>Ex tempore dicere</i> .
 Iſnènade beſjediti . Nènadno govoriti . Rechi nèmiſliuſci . Dir male d' alcuno . <i>Male-dice de aliquo dicere</i> . ſlo o tkomu govoriti . ſlòrechi od tkoga , recem , kaòfam . Koſiti tkoga jeſikom . V. Detrarre . Chi dice mal d' alcuno . <i>Obloquutor , ris . m.</i> ſlòrek , ka . m. ſlògovornik , ka . m. ſlògovoraz , za . m. Il femm. ſlògovorniza , ze .
 Dir all' orecchio . <i>In aure dicere</i> . Poſcjap-tati , ſcjapchjem , taòfam . Scjaptati , chjem , taòfam . Govoriti na uho . Freq. Scjapchi-vati , vam , vòfam . Poſcjapchivati , vam , vòfam . Dire a propoſito . <i>Apte dicere</i> .
 Sputno govoriti . Skladno , podobno , po putu govoriti . Dire a ſpropoſito . <i>Inepte dicere</i> . Nèſputno beſjediti . Nèskladno , nè-podobno govoriti . A dir il vero . <i>Ut ve-rum dicam</i> . ſa rechi pravo . ſa rechi iſti-nu . Dico il vero . <i>Verum dico</i> . Pravvo dim . Govorim iſtinu . Dir di cuore . <i>Ex animo dicere</i> . Govoriti od farza . Govoriti iſ far-za . Dir da dovero . <i>Serio dicere</i> . Govoriti na sbigl . Dire a mente . <i>Dicere memoriter</i> . Rechi na pamet . Dire leggendo . <i>Dicere de ſcripto</i> . Rechi iſ kgnighe . Dir non vo-lendo , dir una parola per un' altra . <i>Labi lingua</i> . Utechi jeſik . ſarechiſe , ricèmſe , ſarekaòſàmſe . Dir l' iſteſſo . <i>Cantilenam candem canere</i> . Pjevati iſtu pjeſan . Sve jedno govoriti . Dir pro , e contra a qual-cuno . <i>In utramque partem dicere</i> . ſa tko-ga , i na tkoga rechi . Dir la coſa come ſta . <i>Rem ut eſt eloqui</i> . Rjeti onako , kakò-je . Dico a te . <i>Tecum loquor</i> . Tebi govo-rim . S' tobom govorim . Queſto ti dico . <i>Hoc ajo</i> . Ovòti velim . Ovo dim .
 <i>Dim Gascpara Milloſtičja</i> ; Ofm. Si dice . <i>Fama eſt</i> . Glas jeſt . Govoriſe . Per dir me-glio . <i>Ut verius dicam</i> . ſa boglje rjeti . Non ſia detto per arroganza . <i>Abſit verbo inju-riæ</i> . Nèbud' receno ſa okolaſt .</p> |
| <p>Diſcépolo . V. Scolaro .</p> | <p>Diſcépolo . V. Scolaro .</p> |
| <p>Diſordinatiſſimo , ed inordinatiſſimo , ſom-mamente diſordinato . <i>Inordinatiſſimus</i> , ma , mum . Prinèredni , na , no . Najnè-rednji , njia , nije .</p> | <p>Diſordinatiſſimo , ed inordinatiſſimo , ſom-mamente diſordinato . <i>Inordinatiſſimus</i> , ma , mum . Prinèredni , na , no . Najnè-rednji , njia , nije .</p> |

| | |
|--|--|
| Diffigillabile , che fi può diffigillare . <i>Resi-
gnabilis</i> , & <i>hoc le</i> . Raspecjatni , na , no .
Odpecjatni , na , no . | Diffigillabile , che fi può diffigillare . <i>Resi-
gnabilis</i> , & <i>hoc le</i> . Raspecjatni , na , no .
Odpecjatni , na , no . |
| Difverginäre . V. Sverginare . | Difverginäre . V. Sverginare . |
| Edile , Magistrato sopra la vettovaglia , e an-
cora sopra glj edificj pubblici . <i>Curator an-
nonæ</i> . Nadbrafcnik , ka . m. Nadljetinik ,
ka . Vifce hrane , illi fgraghje . | Edile , Magistrato sopra la vettovaglia , e an-
cora sopra glj edificj pubblici . <i>Curator an-
nonæ</i> , Nadbrafcnik , ka . m. Nadljetinik ,
ka . Vifce hrane , illi fgraghje . |
| Erbame . <i>Herbæ</i> . f. pl . Travje , vja . n.
----- | Erbame . <i>Herbæ</i> . f. pl . Travje , vja . n. |
| Jestojskàje mâ jedina
<i>Nèpofnano</i> , garko travje . Rafm. | Jestojskàje mâ jedina
<i>Nèpofnano</i> , garko travje . Rafm. |
| Explicare , far intendere . <i>Explico</i> , cas . Ka-
fati , fcem , faòfam . Tomaciti , cim , ciò-
fam . | Explicare , far intendere . <i>Explico</i> , cas . Ka-
fati , fcem , faòfam . Tomaciti , cim , ciò-
fam . |
| Effere , nome , cioè ftato , effenza , efiften-
za . <i>Status</i> , us . Bitje , tja . n. | Effere , nome , cioè ftato , effenza , efiften-
za . <i>Status</i> , us . Bitje , tja . n. |

| | |
|---|---|
| <p>Faccia , volto , parte nota del corpo , <i>Facies</i> , <i>ciei</i> . f. Lize , liza . n. Obraf , fa . m. Faccia tonda . <i>Facies orbiculata</i> . Oblo lize . Faccia puerile . <i>Species puerilis</i> . Djetinski obraf . Faccia di morto . <i>Facies cadaverosa</i> . Martacko lize . Faccia lunga . <i>Facies longa</i> . Dughi obraf . Faccia pallida , e miserabile . <i>Facies miserabilis</i> . Lize bljedo , i tufcno . Proverb. del volgo . <i>Takògàje liza , kò da ni na Boscich nje slana ifio</i> .</p> <p>Defcr .</p> <p><i>fagugnastio</i> , i obrasto
 <i>Mustafàse</i> vas ukafa ;
 Lize suho , bljedo , i tmasto ,
 Pun skoncjagna , pun porafa . Osm.</p> <p>Defcr .</p> <p><i>fagugnastio</i> , i farasto
 Ovi u kosah vasse vidi ;
 Postarano lize , i tmasto ,
 Kafce od kosce suhor blidi . Osm. Faccia spaventevole . <i>Facies horrenda</i> . Obraf strafejan . Faccia scolorita . <i>Facies decolorata</i> . Lize potamgneno . Faccia da galantuomo . <i>Facies liberalis</i> . Lize plemenito . Faccia piena di brofole . <i>Pustulata rubro facies</i> . Obraf puhvicjast . Obraf zargljenim puhvizam pobjen . Faccia brutta . <i>Facies deformis</i> . Gard obraf . Grubo lize . Nèskladno oblicje .</p> <p><i>Pogrescpana</i> , stara liza ,
 <i>Sbabgljen Zarnaz</i> , gard savisce . Osm. A prima faccia . <i>Prima specie</i> . Na parvi pogled . A faccia a faccia . <i>Facie ad faciem</i> . Lizem k' lizu . Lize na lizu . Non ha faccia di parlare . <i>Erubescit loqui</i> . Nèima obras , fa govoriti . Sràmgađe , govoriti . Faccia con rughe . <i>Facies rugosa</i> . Lize pogrescpano .</p> <p><i>Pogrescpana</i> , stara liza ,
 <i>Sbabgljen Zarnaz</i> , gard savisce . Osm. Guardar uno in faccia . <i>Alicujus os intueri</i> . Gledati tkoga u obraf . Gledat tkoga u occhi . <i>fapiscdritle</i> u cigovo lize , alli u tkoga . Gli affari avevano mutata faccia . <i>Magna erat rerum facta commutatio</i> . Velikàje bila promjena u stvarjeh , alli u posljeh . Bilèfu stvarjeh promjenile sliku . Poslisufe bili priobrasili . Uomo senza faccia , vale senza vergogona . <i>Impudens</i> . Govjek bèfobrasan , alliti befocjan .</p> | <p>Faccia , volto , parte nota del corpo . <i>Facies</i> , <i>ciei</i> . f. Lize , liza . n. Obraf , fa . m. Faccia tonda . <i>Facies orbiculata</i> . Oblo lize . Faccia puerile . <i>Species puerilis</i> . Djetinski obraf . Faccia di morto . <i>Facies cadaverosa</i> . Martacko lize . Faccia lunga . <i>Facies longa</i> . Dughi obraf . Faccia pallida , e miserabile . <i>Facies miserabilis</i> . Lize bljedo , tufcno . Proverb. del volgo . <i>Takògàje liza , kò da ni na Boscich nje slana ifio</i> .</p> <p>Defcr .</p> <p><i>fagugnastio</i> , i obrasto
 <i>Mustafàse</i> vas ukafa ;
 Lize suho , bljedo , i tmasto ,
 Pun skoncjagna , pun porafa . Osm.</p> <p>Defcr .</p> <p><i>fagugnastio</i> , i farasto
 Ovi u kosah vasse vidi ;
 Postarano lize , i tmasto ,
 Kafce od kosce suhor blidi . Osm. Faccia spaventevole . <i>Facies horrenda</i> . Obraf strafejan . Faccia scolorita . <i>Facies decolorata</i> . Lize potamgneno . Faccia da galantuomo . <i>Facies liberalis</i> . Lize plemenito . Faccia piena di brofole . <i>Pustulata rubro facies</i> . Obraf puhvicjast . Obraf zargljenim puhvizam pobjen . Faccia brutta . <i>Facies deformis</i> . Gard obraf . Grubo lize . Nèskladno oblicje .</p> <p><i>Pogrescpana</i> , stara liza ,
 <i>Sbabgljen Zarnaz</i> , gard savisce . Osm. A prima faccia . <i>Prima specie</i> . Na parvi pogled . A faccia a faccia . <i>Facie ad faciem</i> . Lizem k' lizu . Lize na lizu . Non ha faccia di parlare . <i>Erubescit loqui</i> . Nèima obras , fa govoriti . Sràmgađe , govoriti . Faccia con rughe . <i>Facies rugosa</i> . Lize pogrescpano .</p> <p><i>Pogrescpana</i> , stara liza ,
 <i>Sbabgljen Zarnaz</i> , gard savisce . Osm. Guardar uno in faccia . <i>Alicujus os intueri</i> . Gledati tkoga u obraf . Gledat tkoga u occhi . <i>fapiscdritle</i> u cigovo lize , alli u tkoga . Gli affari avevano mutata faccia . <i>Magna erat rerum facta commutatio</i> . Velikàje bila promjena u stvarjeh , alli u posljeh . Bilèfu stvarjeh promjenile sliku . Poslisufe bili priobrasili . Uomo senza faccia , vale senza vergogona . <i>Impudens</i> . Govjek bèfobrasan , alliti befocjan .</p> |
| <p>Faldáto , che ha la falda . <i>Fimbriatus</i> , ta , tum . Skutaft , skutafsta , skutafsto .</p> | <p>Faldáto , che ha la falda . <i>Fimbriatus</i> , ta , tum . Skutaft , skutafsta , skutafsto .</p> |
| <p>Fanciullezza . V. Puerilità .</p> | <p>Fanciullezza . V. Puerilità .</p> |
| <p>Faticosissimo . <i>Laboriosissimus</i> , ma , mum . Pri-trudni , dna , dno . Najtrudnji , a , e . Najmucnji , jia , jie .</p> | <p>Faticosissimo . <i>Laboriosissimus</i> , ma , mum . Pri-trudni , dna , dno . Najtrudnji , a , e . Najmucnji , jia , jie .</p> |

| | |
|---|---|
| <p>Fatto , add. <i>Factus</i> , <i>ta</i> , <i>tum</i> . Ucignen , ena , eno . Proverb. <i>Velèje od recenoga do ucignenoga</i> . Dal detto al fatto è un gran tratto . Tvoren. Satvoren , na , no . Dospi-ven , ena , eno . Svarſcen , na , no . Dopo il fatto . <i>Re facta</i> . Po naſada . Po ſto-ru . Po djellu .
 <i>Ersèje pak' ſaman po ſtoru kajati</i> . Ragn. Gljub. Mezzo fatto . <i>Semifactus</i> , <i>ta</i> , <i>tum</i> . Nàpo ucignen. Nàpola ucignen. Polùci-gnen , ena , eno . Non fatto . <i>Infectus</i> , <i>cta</i> , <i>ctum</i> . Nèucignen , ena , eno . Nèſtvo-ren , ena , no . Nèſvarſcen , na , no . Nè-dospiven , ena , eno . Fatto maſtrevolmen-te . <i>Fabrefactus</i> , <i>cta</i> , <i>ctum</i> . Skladnòucignen . Hitroucignen , alli iſhitren , na , no . Fat-to alla groſſa . <i>Pingui minerva factus</i> . O-teſan na debelo . Ucignen na debelo . Fat-to a lumaca . <i>In cochleam retortus</i> . Uci-gnen na ſpuſc , na iſvarte , u ſavitak . Fat-to , cioè ſtagionato , e maturo . <i>Maturus</i> , <i>ra</i> , <i>rum</i> . ſrio , rela , elo . Mezzo ſtagio-nato , mezzo fatto . <i>Semicotus</i> , <i>ta</i> , <i>tum</i> . Polùſrio , rela , lo . Nàpo ſrio . Fatto a ſca-la . <i>Gradatus</i> , <i>ta</i> , <i>tum</i> . Ucignen na liſtve .</p> | <p>Fatto , add. <i>Factus</i> , <i>ta</i> , <i>tum</i> . Ucignen , ena , eno . Proverb. <i>Velèje od recenoga do ucigne-noga</i> . Dal detto al fatto è un gran trat-to . Tvoren. Satvoren , na , no . Dospi-ven , ena , eno . Svarſcen , na , no . Do-po il fatto . <i>Re facta</i> . Po naſada . Po ſto-ru . Po djellu .
 <i>Ersèje pak' ſaman po ſtoru kajati</i> . Ragn. Gljub. Mezzo fatto . <i>Semifactus</i> , <i>ta</i> , <i>tum</i> . Nàpo ucignen . Nàpola ucignen . Polùci-gnen , ena , eno . Non fatto . <i>Infectus</i> , <i>cta</i> , <i>ctum</i> . Nèucignen , ena , eno . Nèſtvo-ren , ena , no . Nèſvarſcen , na , no . Nè-dospiven , ena , eno . Fatto maſtrevolmen-te . <i>Fabrefactus</i> , <i>cta</i> , <i>ctum</i> . Skladnòucignen. Hitroucignen , alli iſhitren , na , no . Fat-to alla groſſa . <i>Pingui minerva factus</i> . O-teſan na debelo . Ucignen na debelo . Fat-to a lumaca . <i>In cochleam retortus</i> . Uci-gnen na ſpuſc , na iſvarte , u ſavitak . Fat-to , cioè ſtagionato , e maturo . <i>Maturus</i> , <i>ra</i> , <i>rum</i> . ſrio , rela , elo . Mezzo ſtagio-nato , mezzo fatto . <i>Semicotus</i> , <i>ta</i> , <i>tum</i> . Polùſrio , rela , lo . Nàpo ſrio . Fatto a ſca-la . <i>Gradatus</i> , <i>ta</i> , <i>tum</i> . Ucignen na liſtve .</p> |
| <p>Fifchiamento , il fiſchiare . <i>Sibili editiæ</i> . ſviſc-dagne , gna . n. ſàſviſcdagne , gna . n. ſviſcd , da . m.</p> | <p>Fifchiamento , il fiſchiare . <i>Sibili editiæ</i> . ſviſc-dagne , gna . n. ſàſviſcdagne , gna . n. ſviſcd , da . m.</p> |
| <p>Fondatiſſimo , molto bene fondato . <i>Funda-tiſſimus</i> , <i>a</i> , <i>um</i> . Prem utemeglni , na , no .</p> | <p>Fondatiſſimo , molto bene fondato . <i>Funda-tiſſimus</i> , <i>a</i> , <i>um</i> . Prem utemeglni , na , no .</p> |
| <p>Fortùna , borraſca di tempo . V. Tempeſta . Fortuna , cioè forte . <i>Fortuna</i> , <i>na</i> . f. Sre-chja , chje . f.
 <i>Tako srechja svim puſiva</i>
 <i>U promjeni vjeclnjeh dobaa</i>
 <i>Promjenjivat svegh uſciva</i>
 <i>Roba u Kraglja , Kraglja u roba</i> . Ofm. Farſi la fortuna . <i>Felicitatem ſibi parare</i> . Graditi ſebi frechju . Sgotoviti frechju . Ca-dere in baſſa fortuna . <i>Labafco</i> , <i>ſcis</i> . Do-chi na magne . Paſtit u potrebu , &c. V. Impoverire . Metter una coſa alla fortu-na . <i>Aliquid fortunæ committere</i> . Spuſtiti na frechju . Fortuna felice . V. Proſperità . Fortuna cattiva . V. Avverſità . Per mala fortuna . <i>Fato infelici</i> . Po hudoj frechi . Per fortuna . V. Caſualmente .</p> | <p>Fortuna , borraſca di tempo . V. Tempeſta . Fortuna , cioè forte . <i>Fortuna</i> , <i>na</i> . f. Sre-shja , chje . f.
 <i>Tako srechja svim puſiva</i>
 <i>U promjeni vjeclnjeh dobaa</i>
 <i>Promjenjivat svegh uſciva</i>
 <i>Roba u Kraglja , Kraglja u roba</i> . Ofm. Farſi la fortuna . <i>Felicitatem ſibi parare</i> . Graditi ſebi frechju . Sgotoviti frechju . Ca-dere in baſſa fortuna . <i>Labafco</i> , <i>ſcis</i> . Do-chi na magne . Paſtit u potrebu , &c. V. Impoverire . Metter una coſa alla fortu-na . <i>Aliquid fortunæ committere</i> . Spuſtiti na frechju . Fortuna felice . V. Proſperità . Fortuna cattiva . V. Avverſità . Per mala fortuna . <i>Fato infelici</i> . Po hudoj frechi . Per fortuna . V. Caſualmente .</p> |
| <p>Frecciàre , e faettare , tirrar frecce , o ferir con frecce . <i>Sagittas emittere</i> . Strjegljati , gljam , gljaòſam . Strjeliti , lim , liòſam . Strjeglivati .
 <i>Jedan poghled bludno strigljen</i>
 <i>Sladak poſmjeh , rjecza od meda , &c.</i>
 <i>Raſm. Strjelivati ſtrjelom . Uſtrjeliti .</i></p> | <p>Frecciàre , e faettare , tirrar frecce , o ferir con frecce . <i>Sagittas emittere</i> . Strjegljati , gljam , gljaòſam . Strjeliti , lim , liòſam . Strjeglivati .
 <i>Jedan poghled bludno strigljen</i>
 <i>Sladak poſmjeh , rjecza od meda , &c.</i>
 <i>Raſm. Strjelivati ſtrjelom . Uſtrjeliti .</i></p> |

Frotta , moltitudine , caterva . *Caterva* , væ ,
f. Jatto , tta . m. Mno/c , o/ci . *f.* Ghljud-
stvo , stva . m. Skup , pa . m. Cetta , tte .
f. V. Moltitudine . In frotta , avv. *Confer-*
tim . U jattu . Skuppo . U jedno . U cetti .

Frotta , moltitudine , caterva . *Caterva* , væ .
f. Jatto , tta . m. Mno/c , o/ci . *f.* Ghljud-
stvo , stva . m. Skup , pa . m. Cetta , tte .
f. V. Moltitudine . In frotta , avv. *Confer-*
tim . U jattu . Skuppo . U jedno . U cetti .

12. Popis kratica navedenih u rječniku i njihova tumačenja

12.1. Popis autora i djela korišten kao korpus za izradu rječnika

U ovom je prilogu naveden popis autora i njihovih djela koja je della Bella koristio kao korpus za izradu rječnika i iz kojih je uzimao primjere kako bi ilustrirao upotrebu pojedine riječi. Navedene su i kratice koje su u tekstu rječnika označavale iz kojeg je djela uzet tekst primjera u rječničkoj natuknici.

Della Bella, na talijanskom jeziku, doslovno kaže sljedeće: „Indeks starih i suvremenih autora citiranih u ovom rječniku s objašnjenjem kratica gdje su naznačene, te također mnoge riječi, u kojima senalaze autori, a da nema citata.“²¹⁷ On dakle uz mnoge natuknice donosi primjere, navodi autora (odnosno djelo), ali isto tako u velikom broju natuknica, bez primjera, navodi djelo odnosno autora u kojem je pojedinu riječ pronašao.

U ovome popisu navedeni su za djela 42 autora hrvatski prijevod i talijanski izvornik objašnjenja, te prema potrebi dodatna objašnjenja u bilješkama.

²¹⁷Della Bella, op. cit., str. 11. "Indice Degli Autori Antichi, e Moderni citati in questo Vocabolario colla spiegazione delle Abbreviature, onde sono indicati; e molte voci de' quali Autori si trovano anche senza le Citazioni".

- Alb. Arhidakon Alberti²¹⁸ odnosno Oficij presvete Djevice Arhidakona Albertija. (Arcidiacono Alberti; o fia Offizio della SS. Vergine dell' Arcidiacono Alberti.)
- Ariadn. Arijadna, tragedija, Ivana Franova Gundulića. Citira se izdanje iz Ankone *Marka Salvionija* iz 1633, u 8. (Arianna; Tragedia di Giovanni di Francesco Gondola. Si cita l' Edizione d' Ancona di *Marco Salvioni* del 1663, in 8.)
- Bab., Babul. Različita djela Miha Bunića Babulinova²¹⁹. Rukopis. Tragedija s naslovom Jokasta, prijevod s grčkog. Rukopis. (Compofizioni varie di Michele Babulinna di Bona. MS²²⁰. Tragedia intitolata Jokaſta, traduzione dal Greco. MS.)
- Bet. Er. Bet. Is. Prijevod Heroide, odnosno Pisma Ovidijevih²²¹ Josipa (Joze) Betondića. Rukopis. Prijevod Tragedije s naslovom Krist Sudac oca Stefana Tuccija iz Družbe Isusove. Rukopis. (Traduzione delle Eroidi, ovvero delle Epistole d' Ovidio di Giuseppe Betondi. MS. Traduzione della Tragedia del P. Stefano Tucci della Compagnia di Gesù, che ha per titolo Cristo Giudice. MS.)
- Bosc. Er. Prijevod prve Ovidijeve Heroide Petra Boškovića. Rukopis (Traduzione della prima Eroide d' Ovidio di Pietro Boſcovich. MS.)

²¹⁸Najvjerojatnije **Matija Matulić Alberti** (1561. – 1623.) *Oficij blažene Marije divice*, (1617). http://hr.wikipedia.org/wiki/Matija_Alberti_Matuli%C4%87 (29. 8. 2014.).

²¹⁹**Miho Bunić Babulinov** (1541. – 1617.) je hrvatski renesansni pjesnik i prevoditelj iz Dubrovnika. http://hr.wikipedia.org/wiki/Miho_Buni%C4%87 (29.8. 2014.).

²²⁰Kratice MS znači *manoscritto* odnosno rukopis.

²²¹Autor je **Publije Ovidije Nazon** (Rim, 20. ožujka 43. pr. Kr. – Constanta, 17. po. Kr.) rimski je pjesnik. http://hr.wikipedia.org/wiki/Publije_Ovidije_Nazon (29. 8. 2014.).

- Bun. Rašl., Raš. Pjesni razlike Ivana Vučića Bunića, odnosno Bone: tj. Oci. Rukopis. Ekloge. Rukopis. Eloge i epitafi. Rukopis. (Poefie varie di Giovanni Vuchichja Bunichja, o fia Bona: cioè gl' Ozj. MS. Egloge. MS. Gl' Elogj, e gl' Epitafj. MS.)
- Bur. Različita djela Marina Borešćića ili Burešćića²²²: tj. Svete pjesme i različite druge. Citira se Venecijansko izdanje Marca Ginama iz 1562. (Composizioni varie di Marino Borefcich o Burefcich: cioè Poefie Sacre e varie altre. Si cita l' Edizione di Venezia del 1562.)
- Daršc. Rašl., Raš. Pjesme Marina Držića: tj. dvije drame, jedna s naslovom Abram, i druga (Le Poefie di Marino Daršcich: cioè due Drammi, uno che ha per titolo Abramo, e l' altro la Natività di N. S. MS. Canzoni ed altre Rime amorose dell' Edizione Veneta di Marco Ginamo del 1630.)
- Dim. Rašl. Dim. Posl. Različita djela Nikole Dimitrijevića: to je pučka prerada psalama pokornih kralja Davida. Citira se izdanje u Veneciji iz 1549. Pjesne duhovne. Rukopis. Izreke izvađene iz antičkih filozofa i Svetoga pisma. Rukopis. Poslanice odnosno pisma u stihovima. Rukopis.²²³ (Opere varie di Nicolò Dimitri: cioè il volgarizzamento dei Salmi Penitenziali di Davidde. Si cita l' Edizione di Venezia del 1549. Poefie Sacre. MS. Sentenze estrate dagli antichi

²²² **Marin Buresić (Burešić)** (Dubrovnik, oko 1510 – Dubrovnik, 1572.) hrvatskijepjesnik.<http://www.enciklopedija.hr/Natuknica.aspx?ID=10229> (29.8. 2014.).

²²³ Nikola Dimitrijević (Dubrovnik, oko 1510. – Kreta 1553. – 1555.): "*Sedam salam pokornijeh kralja Davida*", Mletci, 1549. (Razlike pjesni duhovne – 14 pjesama izdanih zajedno sa zbirkom); Poslanice, 4 poslanice upućene Nikoli Nalješkoviću, iz Stolnog Biograda 1546. godine, iz Aleksandrije 1553. godine; Pričice uzete iz Svetog pisma (niz mudrih rečenica pretočenih u stihove) http://hr.wikipedia.org/wiki/Nikola_Dimitrović (29. 8. 2014.).

Filofofi e dalla Sacra Scrittura. MS. Poslanize, ovvero Lettere in verfi. MS.)

- Ekt. Ekt. Rib. Različite pjesme Petra Hektorovića. Najviše se citira rukopis Ribagne Petra Ektorovichja²²⁴ odnosno na tal. Pescagione di Pietro Ektorovich. Rukopis. (Poefie varie di Pietro Ektorovich. Per lo più fi cita il MS. Ribagne Petra Ektorovichja, o fia la Pescagione di Pietro Ektorovich. MS.)
- Elekt., Elek. Elektra, Sofoklova²²⁵ tragedija. Prijevod s grčkoga Dominka Zlatarića. Citira se venecijansko izdanje Marca Ginama iz 1621. u 16. (Elettra; Tragedia di Sofocle. Traduzione dal Greco di Domenico Šlatarich. Si cita l' Edizione di Venezia di Marco Ginamo del 1621. in 16.)
- Gers. Kass. Prijevod djela Nasljedovanje Krista Tommasa Kempisa koji je načinio Bartul Kašić. Citira se tisak u Rimu iz 1641. u 12. (Traduzione dell' Imitazione di Cristo di Tommaso de Kempis fatta da Bartolomeo Kaffich. Si cita la Stampa di Roma del 1641. in 12.)²²⁶
- Gjor., Gjore. Različite pjesme Gjore, odnosno Giorgia Držića. Rukopis. (Poefie varie di Gjore, ovvero Giorgio Darščich. MS.)

²²⁴Petar Hektorović (Hvar, 1487.–Hvar, 13. ožujka 1572.), hrvatski renesansni pjesnik i rudar. *Ribanje i ribarsko prigovaranje* najvažnije je njegovo djelo dovršeno 14. siječnja 1556. godine, a objavljeno je 1568. u Veneciji (Mlecima). http://hr.wikipedia.org/wiki/Petar_Hektorović (29. 8. 2014.).

²²⁵**Sofoklo** (Kolofon, 495. pr. Kr. – Atena, 406. pr. Kr.) starogrčki je dramatičar. <http://hr.wikipedia.org/wiki/Sofoklo> (29. 8. 2014.).

²²⁶Bartul Kašić, *Tomme od Kempisa kanonika regulara od reda s. Agustina Piismo od nasledovanya Gospodinna Naffcega Yeffufa*, Rim, 1641. http://hr.wikipedia.org/wiki/Bartol_Kašić (31. 8. 2014.).

- Gjorg. Ps., Gjor. Ps. Gjorg. Rašl., Gjor. Raš. Gjorg. S. Ben., Gjor. S. Ben. Gjorg. Dav. Gjorg. Salm. Nad., Gjor. Sal. Nad. Poetske parafraze prema Psalmima Davidovim oca Ignjata Đurđevića benediktinskog monaha. Tiskane u Veneciji od *Kristofora Zane* 1729 u 4. velika. Različiti spjevovi istoga, tj. ljubavne i moralne pjesme, ekloge, događaji, kancone, i prijevod. Sve u rukopisu. Život sv. Benedikta. Taj se u prvom venecijanskom izdanju citira prema rukopisu pisanom perom, a u ovom izdanju obrađuje se izdanje tiskano u Dubrovniku 1784. godine u 4. Život kralja Davida, koje je tiskano prije gore spomenute parafraze i psalama. Bilješke ili komentare na Psalme iste tiskovine u Veneciji od Kristofora Zane iz 1729 koje su tiskane s gore rečenim parafrazama. (Parafrasi poetiche sopra i Salmi di David del P. Ignazio Gjorgi Monaco Benedettino. Stampate in Venezia per *Cristoforo Zane* nel 1729. in 4. grande. Poesie diverse del medesimo, cioè Poesie Amoroſe e Morali, Egloge, Avvenimenti, Canzoni, e Traduzioni. Il tutto MS. Vita di S. Benedetto. Questa nella prima Edizione Veneta ſi citava ſcritta a penna, nella prefente ſi è adoperata l' Edizione ſtampata in Ragufa del 1784. in 4. Vita del Re Davide, che va impreſſa avanti le Parafrasi sopra i Salmi. Note, o ſia Comento sopra i Salmi della medefima Stampa di Venezia di *Cristoforo Zane* del 1729 che vanno impreſſe con le ſuddette Parafrasi.)²²⁷

²²⁷ **Ignjat Đurđević** poznat kao **Ignazio Giorgi** (Dubrovnik, 13. veljače 1675. – Dubrovnik, 21. siječnja 1737.), hrvatski barokni pjesnik i prevoditelj, povjesničar, astronom, biograf. Đurđevićev prijevod ili parafraze psalama kralja Davida *Saltijer Slovinski* izdani su 1729. g. u Veneciji. http://hr.wikipedia.org/wiki/Ignjat_Đurđević (31. 8. 2014.). Također je autor pjesama *Pjesni ljuvene*.

- Gljub. Prijevod djela Aminta od Torqatta Tassa²²⁸ koji je načinio Dominko Zlatarić, šumska bajka. Obrađivano je uvijek izdanje iz Venecije *Domenika i Giovannija Geuerrija* iz 1680. u 8. (Traduzione fatta da Domenico Šlutarich dell' Aminta di Torquatto Taffo, favola boschereccia. Si è adoperata sempre la Stampa di Venezia di *Domenico e Giovanni Guerri* del 1680. in 8.)²²⁹
- Guc. Ođanost Madoninog Rozarija. Djelce Oca Angela Gučetića iz reda Propovjednika. Citira se Venecijansko izdanje Bortola Occhija. (Divozione del Rofario della Madonna: Opufcolo del Padre Angelo Gucetich dell' Ordine de' Predicatori. Si cita l' Edizione di Venezia di Bortolo Occhi.)
- Gund. Rašl., Gun. Raš. Gund. Dubr., Gun. Dubr. Različita djela Ivana Franova Gundulića²³⁰, tj. kratka pjesma u slavu Ferdinanda velikog vojvode od Toskane, *Ljubovnik sramežljiv, Armida*; sve u rukopisu. *Dubravka*, tj. Silvana, šumska bajka. Rukopis. (Compozizioni varie di Giovanni di Francesco Gondola, cioè un Poemetto in lode di Ferdinando Gran-Duca di Toscana, l' Amante Timido, Armida; il tutto MS. *Dubravka*, ovvero Silvia, favola boschereccia. MS.)

²²⁸**Torquato Tasso** (Sorrento, 11. ožujka 1544. – Rim, 25. travnja 1595.) talijanski je pjesnik. <http://www.enciklopedija.hr/Natuknica.aspx?ID=60511> (31.8. 2014.).

²²⁹**Dominko Zlatarić** (Dubrovnik, 1558. –1613.), *Ljubmir, pripovijes pastirska*, Venecija, 1597. http://hr.wikipedia.org/wiki/Dominko_Zlatarić (31.8. 2014.).

²³⁰**Ivan Gundulić** (Dubrovnik, 8. siječnja 1589. – Dubrovnik, 8. prosinca 1638.) hrvatski je barokni pjesnik, epik, lirik i dramatik. http://hr.wikipedia.org/wiki/Ivan_Gundulić (31. 8. 2014.).

- Jegh. Jeghjupka, *Jeđupka* odnosno na tal. *Zingara*. Kratka pjesma Andrije Čubranovića. Citira se izdanje u Veneciji iz 1599. (o fia la Zingara: Poemetto di Andrea Gjubranovich. Si cita l' Edizione di Venezia del 1599.)²³¹
- Ivann., Ivan. Različita djela Ivana Ivaniševića²³². Citira se najviše izdanje iz Venecije. (Composizioni varie di Giovanni Ivanniscevich. Si cita per lo più l' Impressione di Venezia.)
- Kass. Vang. Kass. Is. Kass. B. M. Kass. Djev. Prijevod koji Evanđelja i Apostolskih Poslanica koji je načinio otac Bartul Kašić²³³ iz Družbe Isusove za sve blagdane i nedjelje u godini. Obrađen je otisak iz Rima *Bernardina Tanija* iz 1641 na papiru. Od istoga Život Isusa Krista. Citira se izdanje iz Rima Svete kongregacije za propagandu vjere iz 1638. u dvanaestini. Život Madone istoga tiskara iz Rima iz 1638, u dvanaestini. Životi djevice istoga autora i istoga tiska. (Traduzione fatta dal P. Bartolomeo Kaffich della Compagnia di Gesù degli Evangelj e delle Epistole degli Apostoli per tutte le Feste e Domeniche dell' anno. Si è adoperata la Stampa di Roma di *Bernardino Tani* del 1641. in foglio. Vita di Gesù Cristo del medesimo. Si cita l' Edizione di Roma della S. Congreg. de Propag.

²³¹**Čubranović, Andrija**, hrv.pjesnik (Dubrovnik, prva pol. XVI. st.). Pripisivane su mu maskerata *Jeđupka*, za koju je poslije utvrđeno da ju je napisao Hvaranin Mikša Pelegrinović, te 44 ljubavne pjesme iz zbornika *Nikše Ranjine*, koje potpisuje Andrija Zlatar, za kojega je također ustanovljeno da nije identičan s A. Čubranovićem. <http://proleksis.lzmk.hr/16578/> (1.9. 2014.).

²³²**Ivan Ivanišević** (Postira na Braču, 1608. – Povelja, 1665.) hrvatski je pjesnik. <http://www.enciklopedija.hr/Natuknica.aspx?ID=28168> (1.9. 2014.).

²³³**Bartul Kašić**, isto *Bartol Kašić*, *Bartholomaeus Cassius*, *Bartolomeo Cassio*, kadšto potpisivan i *sBogdančići*/ili s dodatkom *Pažanin*; (Pag, 15. kolovoza 1575.–Rim, 28. prosinca 1650.), hrvatski pisac, jezikoslovac i isusovac. http://hr.wikipedia.org/wiki/Bartol_Kašić (3.9. 2014.).

Fide del 1638. in 12. Vita della Madonna della medesima Stampa di Roma del 1638. in 12. Vite delle Vergini dello stesso Autore e della medesima impressione.)

- Kom. Ogledalo, o sakramentalnoj ispovijedi, Aleksandra Komulovića.²³⁴ Tiskano. (Specchio, della Confessione Sacramentale di Alessandro Komuli. Stampata.)
- Luk. Neka djela Frana Lukarevića Burine,²³⁵ tj. Pjesni razlike. Rukopis. Njegov prijevod s grčkoga tragedije s naslovom Atmante. Rukopis. (Alcune Opere di Francesco Lukari Burina: cioè Poesie diverse. MS. Traduzione da lui fatta dal Greco della Tragedia intitolata Atamante. MS.)
- Mand. Bun. Magdalena pokajnica. Pjesmica Ivana (Džive) Sarova Vučića-Bunića.²³⁶ Citira se otisak iz Venecija *Bortola Occhija* iz 1705. u dvanaestini. (Madalena Penitente; Poemetto di Giovanni di Serafino Vucich-Bona. Si cita l' Impression di Venezia di *Bortolo Occhi* del 1705. in 12.)

²³⁴**Komulović, Aleksandar**, hrvatski crkveni diplomat i nabožni pisac (Split, 1548 – Dubrovnik, 11.VI. ili 2.VII.1608). Glavna djela: *Nauk karstjanski za slovinjski narod* (1582), katekizam napisan na štokavskom narječju i ikavskom govoru, i *Nauk karstjanski kratak* (1603), prijevod katekizma R. Bellarmina. <http://www.enciklopedija.hr/Natuknica.aspx?ID=32681> (3.9. 2014.).

²³⁵**Fran Lukarević Burina** (1541.—1598.) bio je hrvatski renesansni pjesnik i prevoditelj iz Dubrovnika. Izplemičke je obitelji Lukarevića. Pisao je na hrvatskom i talijanskom. Sačuvan esumusamo četiri pjesme koje su prijevod tragedije Girolama Zoppijai Guarinija. http://sh.wikipedia.org/wiki/Fran_Lukarević (3.9. 2014.).

²³⁶**Đivo Sarov Bunić** (Talijanski: *Giovanni Serafino Bona*; 1591. ili 1592. – 6. ožujka 1658.), danas uglavnom poznat kao **Ivan Bunić Vučić**, bio je hrvatski političar i pjesnik iz Dubrovačke republike. http://en.wikipedia.org/wiki/Ivan_Bunić_Vučić (3.9. 2014.).

- Mand. Gjorg., Gjor. Uzdasi Mandalijene pokornice, pjesma oca Ignjata Đurđevića, benediktinskog opata i monaha. Citira se izdanje iz Venecije *Kristofora Zane* iz 1728. u velikoj četvrtine. (I Sospiri di Madalena; Poema del Padre Ignazio Gjorgi Abate e Monaco Benedettino. Si cita l' Edizione di Venezia di *Cristofaro Zane* del 1728. in 4. grande.)
- Mašc. Različite pjesme Horacija²³⁷ i Marina (Maroja) Mažibradića²³⁸ zvanih Šuljaga. Rukopis. (Poefie varie di Orazio e Marino Mašcibratic fopranominati Scjugljaga. MS.)
- Minc., Min. Trubglja Slovinska, ili Trublja Ilirska Vladislava Jeronimovog Menčetića.²³⁹ Korišteno je izdanje iz Ankone, 1665 godine. (o fia Tromba illirica di Vladislavo di Girolamo Menze, ovvero Mincetich. Si è fatto ufo della Edizione d' Ancona del 1665.)
- Nal. Nal. Kgn. Nal. Kom. Različite pjesme Nikola Nale odnosno Nalješkovića.²⁴⁰ Rime ljubavne, moralne i svete. Rukopis. Poslanice 57. pitanja i odgovora u stihovima, rukopis. Komedije 7. u stihovima. Rukopis. (Poefie varie di Nicolò Nale, ovvero

²³⁷**Mažibradić, Horacije (Oracio)**, hrvatski pjesnik (Dubrovnik, oko 1570 – Cavtat, 21.X.1639). Sin pjesnika Maroja Mažibradića. <http://www.enciklopedija.hr/Natuknica.aspx?ID=39651> (3. 9. 2014.).

²³⁸**Maroje Mažibradić Šuljaga** (1519.– 1591.), bio je hrvatski renesansni pjesnik iz Dubrovnika. http://hr.wikipedia.org/wiki/Maroje_Mažibradić (3.9. 2014.).

²³⁹**Vladislav** (Jeronimov) **Menčetić** (Minčetić) (Dubrovnik, oko 1617.–Dubrovnik, 1666.) bio je hrvatski pjesnik iz Dubrovnika. http://hr.wikipedia.org/wiki/Vladislav_Menčetić (3.9. 2014.).

²⁴⁰**Nikola Nalješković** (Dubrovnik, oko 1500.– Dubrovnik, 1587.) hrvatski pjesnik, dramatik i znanstvenik. http://hr.wikipedia.org/wiki/Nikola_Nalješković (3. 9. 2014.).

Naljeckovich: cioè Rime Amoroſe, Morali, e Sacre. MS. Epiſtole 57. di propoſta e di riſpoſta in verſi; MS. Commedie 7. in verſi. MS.)

- Osm. Osman. Herojski spjev Ivana Franova Gundulića, koji ſadrži dvadeset pjevanja. Citira ſe ſtalno tekst piſan perom, iako ſe joſh uvijek očekuje dobro izdanje. (Oſmanide: Poema Eroico di Giovanni di Francesco Gondola, che contiene venti Canti. Si cita ſempre il Teſto a penna, giacchè ancora ſi deſidera una buona Edizione.)
- Palm. Chr. Palm. Zapt., Pal. Zap. Palm. Dan. Palm. Pavl., Pal. Pav. Prijevod Pjesme o životu Isusa Krista, Girolama Vide koji je načinio Junije Palmotić.²⁴¹ Citira ſe otisak iz Rima kod *Giacoma Mascardija*, iz 1670. u 4. gr. *Zaptislava*, odnosno Drama o osnutku staroga Epidaura, kolonije starih Rimljana. Rukopis. Drama koja ima naslov Pavlimir, odnosno (na tal.) Pavao Mirotvorac, u kojoj ſe prikazuje osnutak Dubrovnika. Citira ſe tekst piſan perom. Drama, nazvana *Danica* odnosno *Zora*. Obrađuje ſe tekst piſan perom. (Traduzione del Poema della vita di Geſù Criſto di Girolamo Vida fatta da Giugno Palmotta. Si cita la Stampa di Roma di *Giacomo Mascardi* del 1670 in 4. gr. *Zaptislava*, ovvero Dramma fu la Fondazione dell' antico Epidavro, Colonia degl' antichi Romani. MS. Il Dramma che ha per titolo Pavlìmir, o ſia Paolo Pacificatore, in cui ſi rappresenta la Fondazione di Ragufa. Si cita il Teſto a penna. Il Dramma che viene intitolato *Daniza*, ovvero l' Auro- ra. Si allega il Teſto a penna.)

²⁴¹ **JunijeDžonoPalmotić**(Dubrovnik,1607.–

Dubrovnik,1657.),hrvatskibaroknikipjesnikidramatičar.http://hr.wikipedia.org/wiki/Junije_Palmotić(3.9.2014.).

- Pjesn. Pok., Pjes. Pok. Prijevod pokornih psalama Davidovih (Pjesmi pokorne kralja Davida) koji je načinio Ivan Franov Gundulić. Koristilo se izdanje *Marca Ginama* iz Venecije iz 1620. u dvanestini. (Traduzione dei Salmi Penitenziali di David fatta da Giovanni di Francesco Gondola. Si è fatto ufo della Edizione di *Marco Ginamo* di Venezia del 1620. in 12.)
- Prim. Eurid. Euridika, Drama Paskoja Primovića-Latinčića.²⁴² Obradeno je izdanje iz Venecija kod *Giovannija Salisa* iz 1617. u osminama. (Euridice Dramma di Pasquale Primovich-Latinicich. Si è adoperata l' Edizione di Venezia di *Giovanni Salis* del 1617. in 8.)
- Ragn. Gljub., Rag. Gl. Ragn. Rašl., Rag. Raš. Ljubavne pjesme Dinka Ranjine.²⁴³ Ove se mogu naći odvojeni u drugim izdanjima iz Venecija kod *Marca Ginama* iz 1632. u šesnaestinama. Pjesme različite, odnosno burleskne, moralne i svete. Citirano je izdanje koje je načinio *Lorenzo Torrentino* u Firenci, 1563. u osminama. (Poefie Amoroſe di Domenico Ragnina. Queste ſi trovano ſeparate dalle altre nell' Edizione di Venezia di *Marco Ginamo* del 1632. in 16. Poefie varie, cioè Burleſche, Morali, e Sacre. Vien citata l' Edizione fatta da *Lorenzo Torrentino* in Firenze nel 1563. in 8.)
- Rašm., Raš. Suze sina razmetnoga. Pjesmica Ivana Franova Gundulića. Citiran je otisak koji je načinio *Bortolo Occhi* iz Venecije, iz 1703. u

²⁴²**Primović(Primojević)Latinčić,Paskoje(Paško)**,hrvatskiknjževnikiprevoditelj(Dubrovnik,1565.–Dubrovnik,1619.).<http://www.enciklopedija.hr/Natuknica.aspx?ID=50406>(3. 9. 2014.).

²⁴³**DinkoRanjina**(Dubrovnik,1536.–Dubrovnik,1607.),hrvatskijerenesansnipjesnikizDubrovnika.http://hr.wikipedia.org/wiki/Dinko_Ranjina(3.9.2014.).

dvanaestini. (Le Lagrime del Figliuol Prodigio; Poemetto di Giovani di Francesco Gondola. Vien citata la Stampa fatta da *Bortolo Occhi* in Venezia nel 1703. in 12.)

- **Rit.** Prijevod Rimskoga rituala koji je načinio otac Bartol Kašić iz Družbe Isusove. Citirano je izdanje datino u Rimu. (Traduzione del Rituale Romano fatta dal P. Bartolomeo Kaffich della Compagnia di Gesù. Vien citata l' Edizione della data di Roma.)
- **Ros. Is.** Život Isusa Krista i Djevice Marije, opata Stipana Ružića. Obraden je otisak iz Venecija, kod *Simonea Occhija* iz 1764, u 4. velike. (Vita di Gesù Cristo e della Madonna dell' Abate Stefano Rofa. Si è adoperata la Stampa di Venezia. di *Simone Occhi* del 1764. in 4. grande.)
- **Scim. Šlat.** Nekoja djelca Šimuna Zlatarića.²⁴⁴ Citirano je izdanje iz Venecije kod *Bortola Occhija*. (Alcuni Opuscoli di Simeone Šlatarich. Vien citata l' Edizione di Venezia presso *Bortolo Occhi*.)
- **Scisc., o Scis.** Šest knjiga Ljubavnih i moralnih pjesama Šiška Vlahova Menčetića,²⁴⁵ odnosno Minčetića zvanog Vlahović. Obraden je tekst pisan perom. (Libri fei di Poefie Amorose e Morali di Sigismondo di Biagio Menze, ovvero Mincetich soprannominato Vlahovich. Si allega il Testo a penna.)

²⁴⁴**Zlatarić, Šimun**, hrvatskipjesnikiprevoditelj (Dubrovnik, oko 1593.–Dubrovnik, 1620.). SinDominka Zlatarića. <http://www.enciklopedija.hr/Natuknica.aspx?ID=67284> (3.9.2014.).

²⁴⁵**Šišmundo (Šiško) Menčetić Vlahović** (1457.–1527.) dubrovačkije vlastelin i hrvatskirenesansnipjesnikkoji je stvarao uglavnom uXV. stoljeću. http://hr.wikipedia.org/wiki/Šiško_Menčetić (3.9.2014.).

- **Šlat. Rašl.** Pjesni razlike Dominka Zlatarića,²⁴⁶ odnosno Ljubavne Rime. Citiran je rukopis Moralne Rime. Obrađen je tekst pisan perom. Epitafi. Obrađen je rukopis. (Poefie varie di Domenico Šlatarich, cioè Rime Amoroſe; vien citato il MS. Rime Morali, viene allegato il Teſto a penna. Gl' Epitafj; fi è adoperato il MS.)
- **Sorg. Mer.** Prijevod Tragedije Maffei²⁴⁷, nazvana Meropa koji je načinio Frano Petrov Sorkočević. Korišten je tekst pisan perom. (Traduzione fatta da Francesco di Pietro Sorgo della Tragedia del Maffei intitolata Merope. Si è fatto uſo del Teſto a penna.)
- **Tir.** Tirena, pastirska bajka Marina Držića.²⁴⁸ Korišteno je ponajviše izdanje iz Venecije Francesca Barilletta iz 1607. te tu i tamo ono *Marca Ginama* priređena u Veneciji 1630 g., u dvanaestini. (Tirena; favola boſchereccia di Marino Darſcich. Si è fatto uſo per lo più dell' Edizione di Venezia di *Francesco Bariletto* del 1607. ed anche delle volte di quella di Marco Ginamo fatta pure in Venezia nel 1630. in 12.)

²⁴⁶**Dominko Zlatarić** (Dubrovnik, 1558. – Dubrovnik, 1613.) hrvatski renesansni pjesnik i prevoditelj. http://hr.wikipedia.org/wiki/Dominko_Zlatarić (3.9.2014.).

²⁴⁷**Scipione Maffei** (Verona, 1. lipnja 1675. – Verona, 11. veljače 1755.) talijanski je književnik. <http://www.enciklopedija.hr/Natuknica.aspx?ID=37962> (3.9.2014.).

²⁴⁸**Marin Držić**, (zvan Vidra), hrvatski pjesnik, dramski pisac, autor političkih tekstova i glumac (Dubrovnik, 1508. – Venecija, 2. svibnja 1567.). <http://www.lzmk.hr/hr/izdanja/natuknice/120-hrvatska-enciklopedija/1055-drzic-marin> (3.9.2014.).

- Vetr. Rašl. Vetr. Abr., Vet. Abr. Šest knjiga različitih pjesama gospodina Mavra Vetranovića Čavčića,²⁴⁹ benediktinskog opata i monaha, tj. Rime svete, znanstvene, moralne i komične. Rukopis. Pelegrin (hodočasnik). Komična herojska pjesmica. Rukopis. Abraham, tj. sveta žrtva Abrahamova. Drama od istoga pisca. Citira se rukopis. (Libri fei di diverse Poefie di D. Mauro Vetrani Cjavcich Abate e Monaco Benedettino: cioè Rime Sacre, Scientifiche, Morali, e Burlesche. MS. Il Pelegrino; Poemetto Eroico Comico. MS. Abramo, cioè il Sacrificio d' Abramo; Drama del medesimo. Si cita il MS.)
- Zork., o Zor. Uzdasi jednog pastira, odnosno Idila Vladislava Jeronimova Menčetića. Rukopis nosi naslov Zorka. (I sospiri di un pastore; ovvero Idillio di Vladislavo di Girolamo Menze. MS. e porta per titolo Zorka.)

²⁴⁹**Mavro Vetranović Čavčić**(Dubrovnik,1482. ili1483. –Dubrovnik,1576.)hrvatskijebenediktinaci književnik, jedan od najplodnijih i najdugovječnijih pjesnika starije hrvatske književnosti. http://hr.wikipedia.org/wiki/Mavro_Vetranovi%C4%87 (3. 9. 2014.).

12.2. Popis ostalih kratica

Tablica pojedinih kratica koje se nalaze u della Bellinom rječniku (Tavola d' alcune altre abbreviature, che fi trovano in queft' Opera).

| Kratica | Talijanski | Hrvatski |
|----------------------|-------------------------------------|-----------------------|
| Add. | <i>Addiettivo.</i> | Pridjev. |
| Arcip. | <i>Arcipelago.</i> | Otočje, arhipelag. |
| Avv., o Avverb. | <i>Avverbio, o Avverbialmente .</i> | Prilog, ili priloški. |
| Capit. | <i>Capitale.</i> | Glavni grad. |
| Caft. - | <i>Castello.</i> | Dvorac. |
| Cit. - | <i>Città.</i> | Grad. |
| Cit. ant. | <i>Città antica.</i> | Drevni grad. |
| Comp. | <i>Comparativo.</i> | Komparativ. |
| Dat. | <i>Dativo.</i> | Dativ. |
| Dim. | <i>Diminutivo.</i> | Umanjenica. |
| Eur. | <i>Europa.</i> | Europa. |
| Europ. | <i>Europea.</i> | Europski. |
| F., o Femm. | <i>Femmenino.</i> | Ženski rod. |
| F. pl. | <i>Femmenino plurále.</i> | Ženski rod množina. |
| Fium. | <i>Fiume.</i> | Rijeka. |
| Freq. | <i>Frequentativo.</i> | Učestali. |
| Ill., o Illir. barb. | <i>Illirico barbaro.</i> | Ilirski barbarski. |
| Imp. | <i>Imperio.</i> | Imperij. |
| Indecl. | <i>Indeclinabile.</i> | Nepromjenljivica. |
| Inf. | <i>Inferiore.</i> | Donji. |
| If. | <i>Ifola.</i> | Otok. |
| M. | <i>Mascolino.</i> | Muški rod. |
| M. pl. | <i>Mascolino plurále.</i> | Muški rod množina. |

| Kratica | Talijanski | Hrvatski |
|-------------------|--|---------------------------|
| Metaf. | <i>Metaforicamente.</i> | Metaforički. |
| Metr. | <i>Metropoli.</i> | Metropola. |
| Mont. | <i>Montagna, o Monte.</i> | Planine ili gore. |
| N. | <i>Neutro.</i> | Srednji rod. |
| N. pl. | <i>Neutro plurale.</i> | Srednji rod množina. |
| Part. | <i>Participio.</i> | Particip. |
| Prepof. | <i>Preposizione.</i> | Prijedlog. |
| Prov., o Proverb. | <i>Proverbio, o Proverbialmente.</i> | Poslovica ili poslovično. |
| Provinc., o Prov. | <i>Provincia.</i> | Provincija. |
| Regn. | <i>Regno.</i> | Kraljevstvo. |
| Sett. | <i>Settentrione, o Settentrionale.</i> | Sjever ili sjeverni. |
| Superl. | <i>Superlativo.</i> | Superlativ. |
| Suft. | <i>Sustantivo.</i> | Imenica. |
| T. Bot., o B. | <i>Termine Botanico.</i> | Botanički izraz. |
| T. Eccl., o Ecc. | <i>Termine Ecclesiastico.</i> | Crkveni termin. |
| T. delli Sp. | <i>Termine delli Speciali.</i> | Stručni termin. |
| T. T. | <i>Termine Teologico.</i> | Teološki pojam. |
| V. | <i>Vedi.</i> | Vidi. |
| V. G. | <i>Verbi Grazia.</i> | Radi primjera. |
| V. Baff. | <i>Voce Baffa.</i> | Nenaglašena riječ. |
| V. Gr. | <i>Voce Greca.</i> | Grčka riječ. |
| V. Pol. | <i>Voce Pollaca.</i> | Poljska riječ. |
| V. Ruff. | <i>Voce Ruffa.</i> | Ruska riječ. |

13. Popis tablica

| | |
|---|-----|
| Tablica 3.1 Tablica kratica s prijevodom | 57 |
| Tablica 3.2 Tablica posuđenica | 60 |
| Tablica 3.3 Tablica strukovnog i znanstvenog nazivlja | 61 |
| Tablica 6.1 Distribucija jezičnih oznaka | 171 |
| Tablica 6.2 Distribucija strukturnih oznaka | 173 |
| Tablica 6.3 Točnost označavanja jezika različitim značajkama | 188 |
| Tablica 6.4 Uspješnost konačnog jezičnog klasifikatora | 190 |
| Tablica 6.5 Točnost označavanja strukture različitim značajkama | 192 |
| Tablica 6.6 Uspješnost konačnog strukturnog klasifikatora | 194 |
| Tablica 6.7 Rezultati brzine ručnog označavanja i ispravljanja automatski
označenih oznaka u 60 minuta | 196 |
| Tablica 6.8 Tablica zabuna za jezične oznake | 198 |
| Tablica 6.9 Tablica zabuna za strukturne oznake | 200 |

14. Popis slika

| | |
|--|-----|
| Slika 4.1 Oporuka kćeri Agape Zadarskoga tribuna Dabra 15. srpnja 999. u Zadru | 88 |
| Slika 4.2 Digitalizirana baština Nacionalne i sveučilišne knjižnice poveznice na mrežna tekstualna izdanja Marulićeve Judite | 91 |
| Slika 4.3 Digitalizirana baština Nacionalne i sveučilišne knjižnice slike digitalnog teksta Marulićeve Judite | 92 |
| Slika 4.4 Engleska verzija uvodne stranice Portala hrvatske rječničke baštine.. | 93 |
| Slika 6. 1 Primjer natuknice | 166 |
| Slika 6.2 Krivulja učenja jezičnih i strukturnih oznaka | 202 |

15. Životopis

Petra Bago (matični broj znanstvenika: 314883) rođena je 18. studenoga 1982. godine u Zagrebu. Osnovnu školu pohađala je u Zagrebu, Bernu (Švicarska) i Sofiji (Bugarska). V. gimnaziju prirodoslovno-matematičkog usmjerenja završila je u Zagrebu.

Godine 2003. upisuje Filozofski fakultet, smjer informacijskih znanosti i etnologije i kulturne antropologije. Od 2004. sudjeluje u nastavi kao demonstratorica na Odsjeku za informacijske i komunikacijske znanosti. Kao suautor 2006. godine prima Rektorovu nagradu za studentski znanstveni rad pod naslovom *Računalna priprema, obrada i prikaz sedmojezičnog rječnika Petera Lodereckera povodom 400. godišnjice izdavanja* izrađen pod vodstvom prof. dr. sc. Damira Borasa i Nikole Ljubešića, tadašnjeg znanstvenog novaka na Odsjeku. Diplomirala je 2008. godine na Odsjeku za informacijske i komunikacijske znanosti i Odsjeku za etnologiju i kulturnu antropologiju.

Akadske godine 2008./2009. upisuje poslijediplomski doktorski studij na Odsjeku za informacijske i komunikacijske znanosti. Od 2009. na istom se odsjeku zapošljava kao znanstvena novakinja na projektu *Hrvatska rječnička baština i hrvatski europski identitet* pod vodstvom prof. dr. sc. Damira Borasa. Aktivno sudjeluje u izvođenju nastave na preddiplomskoj razini na kolegijima *Računalne mreže* i *Jezične baze podataka*. Pohađala je više tečajeva, ljetnih škola i europskih intenzivnih programa u zemlji i inozemstvu.

Područja interesa: računalna leksikografija, enciklopedistika, kodiranje teksta, društveno-humanistička informatika, statistička analiza.

15.1 Popis objavljenih radova

15.1.1 Poglavlja u knjizi

1. **Bago, Petra**; Lauc, Tomislava; Boras, Damir. Using quantitative methods for a student activity analysis in an online graduate course concerning their undergraduate education // Recent Researches in Mathematical Methods in Electrical Engineering & Computer Science / Thomas, Gregoire ; Fleaurant, Cyril ; Panagopoulos, Thomas ; Chevassus-Lozza, Emmanuele (ur.). Angers : WSEAS Press, 2011. Str. 122-127.

2. Lauc, Tomislava; **Bago, Petra**; Kišiček, Sanja. Using quantitative methods for a student study activity analysis in a virtual learning environment concerning different students' backgrounds // Computers in Education / Čičin-Šain, Marina ; Uroda, Ivan ; Turčić Prstačić, Ivana ; Sluganović, Ivanka (ur.). Rijeka : Hrvatska udruga za informacijsku i komunikacijsku tehnologiju, elektroniku i mikroelektroniku, 2011. Str. 259-262.

15.1.2 Znanstveni radovi u drugim časopisima

1. Kisicek, Sanja; Boras, Damir; **Bago, Petra**. Designing Educational Contents in and for the Electronic Environment. // *Information technology interfaces (ITI)*. 1 (2010) ; 403-408 (članak, znanstveni).

2. Ljubešić, Nikola; **Bago, Petra**; Boras, Damir. Statistical Machine Translation of Croatian Weather Forecasts: How Much Data Do We Need?. // *CIT. Journal of computing and information technology*. 18 (2010) , 4; 303-308 (članak, znanstveni).

15.1.3 Znanstveni radovi u zbornicima skupova s međunarodnom recenzijom

1. Ljubešić, Nikola; **Bago, Petra**; Boras, Damir. Statistical machine translation of Croatian weather forecast: How much data do we need? // *Proceedings of the ITI 2010 32nd International Conference on INFORMATION TECHNOLOGY INTERFACES* / Luzar-Stiffler, V. (ur.). Zagreb : University Computing Centre, University of Zagreb, 2010. 91 (predavanje, međunarodna recenzija, objavljeni rad, znanstveni).

2. **Bago, Petra**; Boras, Damir; Ljubešić, Nikola. First Steps Toward Developing a System for Terminology Extraction // *INFuture2009: Digital Resources and Knowledge Sharing* / Stančić, Hrvoje ; Seljan, Sanja ; Bawden, David ; Lasić-Lazić, Jadranka ; Slavić, Aida (ur.). Zagreb : Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2009. 197-206 (predavanje, međunarodna recenzija, objavljeni rad, znanstveni).

15.1.4 Sažeci u zbornicima skupova

1. Kišiček, Sanja; Lauc, Tomislava; **Bago, Petra**. Students' Usage and Access to Multimedia Learning Resources in an Online Course with Respect to Individual Learning Styles as Identified by the VARK model // *21st Century Learning for 21st Century Skills (LNCS 7563)* / Ravenscroft, Andrew ; Lindstaedt, Stefanie ; Delgado Kloos, Carlos ; Hernández-Leo, Davinia (ur.). Saarbrücken : Springer, 2012. 548-548 (poster, međunarodna recenzija, sažetak, znanstveni).

15.1.5 Neobjavljena sudjelovanja na skupovima

1. **Bago, Petra**; Boras, Damir. A framework for consolidating most important digitized Croatian dictionaries from 1595 to 1945 // *17. International Information Technology and Journalism conference*. (predavanje, ppt prezentacija).

2. **Bago, Petra**; Boras, Damir; Kišiček, Sanja. Web 2.0 Technologies in Journalism: Are Web 3.0 and Journalism 3.0 in sight? // *15. International Information Technology and Journalism conference*. (predavanje, ppt prezentacija).

3. Kišiček, Sanja; Boras, Damir; **Bago, Petra**. Od mentalne mape do multimedijskog obrazovnog sadržaja // *Znanstveno stručni skup Tiskarstvo 09* (<http://www.ziljak.hr/tiskarstvo/tiskarstvo09/>). 2009. (predavanje, međunarodna recenzija, ppt prezentacija, znanstveni).

4. Ljubešić, Nikola; Boras, Damir; **Bago, Petra**. Language Technology on News Portals // *14. International Information Technology and Journalism conference*. Dubrovnik, 2009. (predavanje, ppt prezentacija).

15.1.6 Druge vrste radova

1. **Bago, Petra**; Kolarec, Martina. Računalna priprema, obrada i prikaz sedmojezičnog rječnika Petera Lodereckera povodom 400. godišnjice izdavanja, 2006. (rektorova nagrada za studentski znanstveni rad).